A. S. KOMPANEYETS

A COURSE OF THEORETICAL PHYSICS

VOLUME 2

STATISTICAL LAWS

STATISTICAL PHYSICS
HYDRODYNAMICS AND GAS DYNAMICS
ELECTRODYNAMICS OF CONTINUOUS MEDIA
PHYSICAL KINETICS

MIR PUBLISHERS · MOSCOW

A. S. KOMPANEYETS

A COURSE OF THEORETICAL PHYSICS

VOLUME 2

ИЗДАТЕЛЬСТВО • МИР •

A. S. KOMPANEYETS

A COURSE OF THEORETICAL PHYSICS

VOLUME 2

MIR PUBLISHERS
MOSCOW

A. S. KOMPANEYETS

STATISTICAL LAWS

Translated from the Russian by V. TALMY

MIR PUBLISHERS
MOSCOW

А. С. КОМПАНЕЕЦ

КУРС ТЕОРЕТИЧЕСКОЙ ФИЗИКИ

ТОМ 2

СТАТИСТИЧЕСКИЕ ЗАКОНЫ

English translation first published 1978
Revised from the 1975 Russian edition

На английском языке

PREFACE

In selecting the material for the second volume of this course of theoretical physics a more subjective approach than in setting forth the elementary laws in the first volume was inevitable. It is natural that in sheer volume the applications surpass the fundamentals. In any case, the material cannot be presented too briefly, because what is not understood is not only wasted on the reader but in addition fosters a feeling of frustration and dislike for the subject.

I attempted to arrange the subject matter in a way that would make it possible to discuss the basic laws from different aspects. This provides a sense of completeness and, according to the golden rule of education, consolidates the body of acquired knowledge.

That is one of the reasons why I have devoted a fair amount of space to gas dynamics, where the most important thermodynamic relationships are involved. Gas dynamics is in itself interesting by virtue of the fact that it explicates special features of nonlinear wave phenomena, such as the appearance of discontinuities in smooth flows, the establishment of steady-state conditions in irreversible processes, and many others. Besides, gas dynamics and hydrodynamics have many applications in modern technology.

Unfortunately there was not space enough in the book for another extremely interesting department of the mechanics of continuous media, the elasticity theory.

The electrodynamics of continuous media is set forth in such a way as to refer more frequently to statistical physics. This should make both these parts of the second volume more clear. Kinetics also includes a section that directly adjoins on statics. The fourth part of the book presents the kinetic equation method and also examines metals and semiconductors. This, of course, is but a small part of physical kinetics, but perhaps the most important.

Here and there some historical notes have been included. Presenting a subject in its development in some cases makes possible a better explanation of the interdependence of, and interconnection between, discoveries, which in theoretical physics have never been works of chance or mere volition.

Like the first volume, the second offers some information of a mathematical nature. In concrete applications they appear much simpler than in special textbooks. It was Enrico Fermi who said: "I

have gained more mathematics from books on physics than from books on mathematics."

In this volume, too, I have frequently referred to the *Course of Theoretical Physics* by L.D. Landau and E.M. Lifshitz. I was greatly assisted by R. Courant and K.O. Friedrichs' book, *Supersonic Flow and Shock Waves*, and two books by G.H. Wannier, *Statistical Physics* and *Elements of Solid Stat Theory*.

I am much indebted to M.I. Kaganov for his advice on the material of Part IV of this volume.

A.S. Kompaneyets

CONTENTS

PART I

STATISTICAL PHYSICS

1

EQUILIBRIUM DISTRIBUTION OF MOLECULES IN IDEAL GAS

The Subject of Statistical Physics. The methods of quantum mechanics set forth in the first volume make it possible, in principle, to describe assemblies of electrons, atoms and molecules comprising macroscopic bodies.

In practice, however, even the problem of an atom with two electrons presents such formidable mathematical difficulties that no one has so far been able to solve it completely. It is all the more impossible not only to solve but even to write the wave equation for a macroscopic body consisting of, say, 10^{23} atoms with their electrons.

However, in large systems we observe certain general laws of motion which can be described without knowing the wave function of the system. Let us give one very simple example of such a law. Assume that there is only one molecule in a large, completely empty vessel. If the molecule's motion is not defined beforehand, the probability of its being in any half of the vessel is equal to 1/2. If there are two molecules, the probability of their being in the same half of the vessel simultaneously is $(1/2)^2 = 1/4$. And the probability of the whole of a gas consisting of N particles being in the same half of the vessel is $(1/2)^N$, that is, an infinitesimally small number. On the average, therefore, there will always be an approximately equal number of molecules in each half of the vessel. The greater the number of molecules that make up the gas the closer to unity the ratio of the numbers of molecules in both halves, whenever they are observed.

This approximate equality of the number of molecules in equal volumes of the same vessel offers an almost obvious example of a statistical law applicable only to large assemblies of identical objects.

In addition to spatial distribution, such an assembly of molecules is also characterized by a certain velocity distribution. Thus, if a gas in a given volume is at rest, there will on the average be the same number of molecules moving in any direction. Less obvious is the distribution of molecules according to their absolute velocities (on this see Sec. 2).

Statistical physics studies the laws governing the motions of large assemblies of electrons, atoms, quanta, molecules, etc. The velocity distribution of molecules is one of the simplest problems solved by the methods of statistical physics.

This department of theoretical physics introduces a number of new quantities, which do not make sense in the dynamics of single bodies or of a small number of bodies. An example of a statistical quantity is *temperature*, which is closely related to the mean energy of a gas molecule. In the statistical approach, averaging is done over a large number of identical bodies. It is important to note that distributions according to various mechanical parameters of a system may occur spontaneously. Thus, if a gas is confined to one-half of a vessel and the partition is then removed, the gas will uniformly fill the whole vessel. Similarly, if the velocity distribution in the gas is in some way disrupted, the initial statistical distribution will be restored as a result of interactions (collisions) between the molecules. Thus, statistical laws derive not only from the involvement of large numbers of objects but also from their interactions.

Statistical Laws in Quantum Mechanics. Quantum mechanics also describes statistical regularities relating, however, to separate objects. They manifest themselves in very large numbers of identical experiments on identical objects and are not concerned with the interactions between them. For example, electrons in a diffraction experiment may pass through a crystal at arbitrary time intervals and nevertheless give exactly the same picture of the blackening of a photographic plate as they would when passing through the crystal at the same time.

Similarly, regularities in alpha decay cannot be traced to the fact that a very large number of nuclei is involved: since in practice the process is not induced by interactions between nuclei, the statistical character of the quantum mechanical predictions is only manifested for a large number of identical objects, but is by no means due to their number. Motion in quantum mechanics is described to an accuracy compatible with the uncertainty principle. Let us now show how to go over to the less exact descriptions of statistical physics.

First, let us suppose that the wave equation for a certain system comprising a very large number of particles has been solved. This corresponds to a precise quantum mechanical description of the system. Let the solution have produced a certain spectrum of the energy

eigenvalues of the system

$$E_0,\ E_1,\ E_2,\ \ldots,\ E_n,\ \ldots \tag{1.1}$$

corresponding to states with wave functions

$$\psi_0,\ \psi_1,\ \psi_2,\ \ldots,\ \psi_n,\ \ldots$$

Then the wave function for any state, as was shown in [Sec. 25][1], can be represented in the form of a sum of wave functions of states with definite energy eigenvalues:

$$\psi = \sum_n c_n \psi_n \tag{1.2}$$

The quantity

$$w_n = |c_n|^2 \tag{1.3}$$

gives the probability that a measurement of the energy of a system in state ψ will yield the nth eigenvalue.

Expansion (1.2) makes it possible to determine not only the amplitudes but also the relative probability phases corresponding to a detailed quantum mechanical description of the system. The methods of statistical physics make it possible to determine approximately the quantities $w_n = |c_n|^2$ without determining the probability phases. Naturally, knowledge of w_n is not sufficient to construct the wave function of the system; however, it is possible to determine the mean values of quantities characterizing macroscopic bodies, for instance, mean energy, which are of practical importance. In this section we will show how to calculate the probability w_n for the case of an ideal gas.

Ideal Gas. The *ideal gas* is a system of particles whose interactions can be neglected. What does this mean? Interactions resulting from collisions between molecules matter only when the statistical distribution w_n is in the process of being established. When it becomes established, the collisions among individual molecules affect it insignificantly and in certain approximations they may be neglected. In such cases we say we have an ideal gas.

In condensed systems, that is, liquids and solid bodies, the molecules are in constant vigorous interaction, so that the statistical distribution is substantially affected by the forces acting between molecules.

But in a gas, too, the particles cannot be regarded as absolutely independent. For example, Pauli's exclusion principle imposes important limitations on the possible states of a gas as a whole: two particles cannot be in exactly the same quantum state. We shall take these limitations into account in calculating probabilities.

[1] References in brackets are to Volume 1 of this course.

The States of Individual Particles in a Gas. To distinguish between the states of separate particles and the states of the gas as a whole we shall denote the energies of the particles by the letter ε and the energy of the gas as a whole by E. Thus, for example, if a gas is contained in a rectangular potential well [Sec. 28], from Eq. [28.19] we obtain the energy values for each particle in the form

$$\varepsilon_s = \frac{\pi^2 h^2}{2m}\left(\frac{s_1^2}{a_1^2} + \frac{s_2^2}{a_2^2} + \frac{s_3^2}{a_3^2}\right)$$

where s_1, s_2, s_3 are positive integers, and a_1, a_2, a_3 are the lengths of the sides of the well.

Let, in the most general case, ε take on the following series of values:

$$\varepsilon_0,\ \varepsilon_1,\ \varepsilon_2,\ \ldots,\ \varepsilon_k,\ \ldots \tag{1.4}$$

If there are n_0 particles in the state with energy ε_0, and in general, there are n_k particles in the state with energy ε_k, the total energy of the gas is

$$E = \sum_k n_k \varepsilon_k \tag{1.5}$$

Because in a system consisting of noninteracting particles the energy is an additive quantity, by setting different combinations of numbers n_k we obtain the energy eigenvalues forming the series (1.1) for the gas as a whole.

Quantum mechanics offers countless examples that the energy ε_k does not uniquely define the state of a system. For example, the energy of a hydrogen atom depends only on the principal quantum number n (not to be confused with the number of molecules n_k), so that at any given energy a hydrogen atom may be in one of the $2n^2$ states [Secs. 29, 30]. The number $2n^2$ is called the *weight of the state with energy* ε_n. However, in principle it is also possible to place a system in such conditions that the energy value defines the state uniquely. Let us first of all note that in all atoms except hydrogen the energy depends not only on the principal quantum number n but also on the orbital angular-momentum quantum number l. Further, account of interaction between electron spin and orbital motion shows that the energy also depends on the total angular-momentum quantum number j. Finally, if the atom is placed in an external magnetic field, the energy also depends on the projection of the angular momentum on the field. Thus, there exist conditions under which the energy fully defines the state of the atom (the splitting of all the $2n^2$ states with the same principal quantum number n).

Going back to the states of particles in a closed vessel, if it has the shape of a box with incommensurable squares of the sides, $(a_1)^2$, $(a_2)^2$, $(a_3)^2$, then any combination of integers s_1, s_2, s_3 yields one and

only one number. If the particles possess an intrinsic angular momentum, we can, so to say, remove the degeneracy by placing the gas in a magnetic field (an energy eigenvalue is said to be *degenerate* if several states of the system correspond to it). We shall first consider only systems with completely removed degeneracy.

The States of an Ideally Closed System. We shall now consider the energy spectrum of a gas consisting of noninteracting particles contained in a closed volume and isolated from external influences. For simplicity, we shall assume that one value of energy corresponds to each state of the system as a whole and, conversely, one state corresponds to one energy value. This assumption is true if *all* the energy eigenvalues of every particle are incommensurable numbers.[2] If we denote these numbers ε_k and if there are n_k particles in the kth state, we find that the total energy is $E = \sum n_k \varepsilon_k$. If E is given with infinite precision, for incommensurable ε_k's it is possible in principle to determine all n_k's from this equation. Note that we are speaking not of determining the state of a separate particle from its energy ε_k but of finding the state of the whole gas from the sum of the energies of the particles. Every interval of values dE, however small (but not infinitely small), will include very many eigenvalues E. And each of them corresponds to its own set of numbers n_k, that is, to a definite state of the system as a whole.

States of a Nonideally Closed System. Energy is an exact integral of motion only in an ideally closed system whose state remains unchanged for an indefinitely long time. Conservation of E provides for the constancy of all n_k's. But there are no, and cannot be any, ideally closed systems in nature. Every system interacts in some way with its surroundings. Let us assume this interaction weak and determine how it affects the behaviour of the system.

Suppose that the interaction with the surroundings does not considerably disturb the quantum levels of the individual particles. Nevertheless, according to the principles of quantum mechanics [Sec. 37], every level ε_k ceases to be a precise number and receives a small but finite width $\Delta\varepsilon_k$. This is sufficient for the meaning of the equation $E = \sum n_k \varepsilon_k$ to change fundamentally: in a system consisting of many particles an equation containing imprecise quantities ε_k no longer defines n_k.

An interaction with the surroundings, no matter how weak, makes a precise determination of the state from the total energy E impossible.

---

[2] In a rectangular box the energy of a state, $\varepsilon(s_1, s_2, s_3)$, is commensurable with $\varepsilon(2s_1, 2s_2, 2s_3)$. Therefore, the energy of all states can be incommensurable only in a box of a more complex shape than rectangular.

Transitions Between Contiguous Energy States. In an ideally closed system transitions between the states in an energy interval dE do not occur because of the law of conservation of energy. In the case of a weak interaction with the surroundings all transitions between different states are possible if they do not change the total energy to an accuracy generally acceptable with the determination of the energy of a nonideally closed system. In other words, transitions are possible in a certain energy interval ΔE given by the relationship

$$\Delta E \, \Delta t \sim 2\pi h$$

where Δt is the time interval during which the system can be treated as a closed one.

Let us now assume that the interaction with the surroundings is so weak that for a small time interval it is possible in principle to determine all the values of n_k and thus state the total energy of the gas $E = \sum n_k \varepsilon_k$. But now over a large interval of time the state of the gas may vary within the limits of the total energy interval due to the inaccuracy in the energies, $\Delta\varepsilon_k$, of separate states. All transitions will occur that are compatible with the approximate equation $E = \sum n_k (\varepsilon_k \pm \Delta\varepsilon_k)$. Obviously, a state in which all $\Delta\varepsilon_k$'s are of one sign is highly improbable, which is why we use the symbol $\pm$. We must find the state that forms as a result of all possible transitions in the interval ΔE.

Probabilities of Direct and Reverse Transitions. A very important relation exists between the probabilities of direct and reverse transitions. Let us first consider it on the basis of formula [32.42] obtained as a first approximation in the perturbation theory. Let a system have two states, A and B, with wave functions ψ_A and ψ_B. To these two states corresponds the same energy value within the limits of the inaccuracy ΔE due to the interaction between the system and the surroundings. Within the interval ΔE both states can be treated as belonging to a continuous spectrum. Then from [32.42] the probability of a transition from A to B per unit time is equal to $2\pi h^{-1} \mid V_{AB} \mid^2 g_B$, and the probability of a transition from B to A is $2\pi h^{-1} \mid V_{BA} \mid^2 g_A$, where

$$V_{AB} = \int \psi_A^* \hat{V} \psi_B \, dx \, dy \, dz$$

$$V_{BA} = \int \psi_B^* \hat{V} \psi_A \, dx \, dy \, dz$$

The symbols g_A, g_B denote the weights of the states. But if $g_A = g_B$, then, as $\mid V_{AB} \mid^2 = \mid V_{BA} \mid^2$, the probabilities of a direct and reverse transition are equal. A transition is possible only by virtue of the fact that the energies E_A and E_B are not defined exactly

and a small interval ΔE is given for which the energy spectrum is continuous (in an ideally closed system $E_A \neq E_B$).

The obtained relationship holds only to a first approximation of the perturbation theory. However, there is also a general relationship, which can be derived from the general principles of quantum mechanics. The form of quantum mechanical equations suggests that in time reversal, that is, when $-t$ is substituted for t, the weight does not change provided a transition from ψ_k to ψ_k^* is simultaneously effected. But it is possible to revert from ψ_k^* to ψ_k provided the signs of all the linear and angular momenta are changed. Hence, the probabilities of transitions from A to B and from B^* to A^* are equal (A^* and B^* differ from A and B in the signs of the linear and angular momenta of the two states).

The Equiprobability of States with the Same Energy. Thus, due to interactions with the surroundings, transitions will occur within a system between all possible states A, B, C, . . . belonging to the same energy interval ΔE. If we wait long enough, the system will spend equal times in the A, B, C, . . . states. This is most easily proved indirectly. Assume, first, that the probabilities of direct and reverse transitions are equal ($W_{AB} = W_{BA}$), and then consider the refined relationship $W_{AB} = W_{B^*A^*}$.

So let $W_{AB} = W_{BA}$. We assume that t_A is greater than t_B, so that the system changes from A to B more frequently than from B to A. But this cannot go on indefinitely as, if the ratio t_A/t_B increases, the system will ultimately occur only in A despite the transitions from A to B. Only the equality $t_A = t_B$ can hold indefinitely (on average) by virtue of the fact that direct and reverse transitions occur on average with equal frequency. The same reasoning suggests that if there are many states for which direct and reverse transitions are equally probable, then, given a sufficiently long period of time, the system will on average spend the same time in each state.

It is natural to assume that $t_{A^*} = t_A$, because the states A and A^* differ only in the signs of all linear and angular momenta (and in the sign of the external magnetic field, which must also be changed so that the magnetic energy of all the particles is the same in states A and A^*). Proceeding from this assumption, we see that all the preceding reasoning can be extended to the more general case of $W_{AB} = W_{B^*A^*}$.

It has thus been shown that the system spends the same time in all states of the same weight belonging to the same interval ΔE.

Probability of an Individual State. We shall call the limit of the ratio t_A/t, when t increases indefinitely, the *probability of the state*, q_A. It follows from the equality of all t_A that the corresponding states are equiprobable. But this makes possible a direct determination

of the probability of each state. Indeed, let p be the number of all states. Then $\sum_{A=1}^{p} t_A = t$ and $\sum_{A=1}^{p} q_A = 1$. And since we have proved that the states are equiprobable, $q_A = 1/p$.

Thus, the problem of determining the probabilities of individual states of an ideal gas is reduced to combinatorial analysis. To make use of its methods we must only additionally determine which states of a system of molecules should be regarded as physically different. In computing the total number p each such state must be taken once.

Determination of the States of an Ideal Gas in Statistics. If a gas consists of identical particles, for example, electrons, helium atoms, etc., its state as a whole is precisely given if we know the number of particles occurring in each of their states. It is meaningless to inquire which particles specifically occur in a certain state since identical particles are in principle indistinguishable. If the spin of the particles is half-integral, *Pauli's exclusion principle* [Sec. 33] holds, and in each state there will occur either one particle or none at all. On this basis we must determine the state of the gas as a whole according to the states of separate particles.

To illustrate the calculation of the number of states of a system as a whole, let us assume that there are only two particles, each of which can occur in only two states, a and b ($\varepsilon_a = \varepsilon_b$), the weight of each state being equal to unity. Leaving Pauli's exclusion principle aside for the moment, we find that only the following three different states of the system are possible:

(a) both particles in state a, state b is unoccupied;
(b) both particles in state b, state a is unoccupied;
(c) one particle in each state.

In view of the indistinguishability of the particles the third state must be counted only once (interchange of identical particles between states is meaningless). If, in addition, the particles are subject to the exclusion principle, only state (c) is possible. Thus, the exclusion principle substantially reduces the number of possible states of a system. If the exclusion principle applies, a system can occur in only one state; otherwise it can occur in three states. A system of two different particles, for instance, an electron and a positron, would have four states.

Let us now consider the example of three identical particles occurring in three states with equal energy. If Pauli's exclusion principle applies, only one state of the system as a whole is possible: one particle occurs in each state. If there is no exclusion the indistinguishable particles can be arranged thus: (1) one in each quantum state; (2) two in one state and the third in one of the two others (which gives six states for the system); and (3) all three particles in any quantum state. Thus we have $1 + 6 + 3 = 10$ states for the system as a whole.

If the three particles were distinguishable (π^+, π^0, π^- mesons, for instance), each could occur in any of the three states independently of the others, and all three particles together would have $3^3 = 27$ states. Later on a general formula for calculating the number of states will be developed. Let us begin with particles with integral or zero spin.

Particles Not Subject to Pauli's Exclusion Principle. For the further deductions there is no need to assume that every state of a particle of given energy has unit weight. We shall denote the weight of a state of a particle of energy ε_k by the symbol g_k. In other words, g_k states of a particle have the same energy ε_k or, more precisely, lie in a certain small energy interval $\Delta\varepsilon_k$ in the neighbourhood of ε_k. These states are equiprobable for all particles.

Let us assume that n_k particles not subject to Pauli's exclusion principle have energy ε_k, and we have to calculate the number of ways these particles can be distributed in g_k states. Let this number be $P(n_k, g_k)$.

As proved before, the probability of each arrangement of the particles by states is $[P(n_k, g_k)]^{-1}$.

In order to calculate $P(n_k, g_k)$ we shall, as is usually done in combinatorial analysis, call the state a "box" and the particle a "ball". The problem is: In how many ways can n_k balls be placed in g_k boxes without identifying the balls, that is, without relevance to which ball lies in which box? If the particles are not subject to Pauli's exclusion principle, each box may contain any number of balls.

Let us mix all the balls and all the boxes so that we obtain $n_k + g_k$ objects. From these we select a box at random and set it aside. There remain $n_k + g_k - 1$ objects, from which we randomly draw one object at a time, irrespective of whether it is a box or a ball, and lay them out in a row from left to right from the first box. The following series may, for example, result:

bx, bl, bl, bx, bx, bl, bl, bl, bx, bl, bx, bl, bl, . . .

Since the first object on the left is, by definition, a box, the remaining objects can be arranged among themselves in $(n_k + g_k - 1)!$ ways.

Now drop each ball into the nearest box to the left. In the series above there will be two balls in the first box, none in the second, three in the third, one in the fourth, etc. The total number of combinations is $(n_k + g_k - 1)!$ but they are not all distinguishable. Indeed, substituting the second ball for the first or any other changes nothing in the series. There are $n_k!$ combinations of the balls. In exactly the same way the boxes can be changed about among themselves since the order in which they occur does not matter. Only the first box cannot be touched as it is there by definition. In all there are

$(g_k - 1)!$ combinations of the boxes. Consequently, of all possible $(n_k + g_k - 1)!$ combinations only

$$P(n_k, g_k) = \frac{(n_k + g_k - 1)!}{n_k!\,(g_k - 1)!} \tag{1.6}$$

are different. If, for example, $n = 3$, $g = 3$, then $P(3, 3) = 5!/(3!2!) = 10$, which is what we obtained before by direct computation.

Particles Subject to Pauli's Exclusion Principle. With particles subject to Pauli's exclusion principle matters are even simpler. Indeed, since no more than one particle can occur in each state, we have the inequality $n_k \leqslant g_k$. Of the total number of g_k states n_k states are occupied.

The number of ways in which we can choose n_k states is equal to the number of combinations of g_k objects n_k at a time:

$$P(n_k, g_k) = C(g_k, n_k) \equiv \frac{g_k!}{n_k!\,(g_k - n_k)!} \tag{1.7}$$

That is the number of possible states of the system when $n_k \leqslant g_k$ and when not more than one particle occurs in any of the g_k states of an individual particle.

The Most Probable Distribution of Particles by State. The numbers g_k and n_k refer to a definite energy. The total number of states of a gas, P, is equal to the product of the numbers $P(n_k, g_k)$

$$P = \prod_k P(n_k, g_k) \tag{1.8}$$

So far we have been employing only combinatorial analysis. It has also been shown that all individual states are equally probable. The quantity P depends on the distribution of particles by state. It can be observed that a gas is in fact always in a state close to the one in which the distribution of individual particles corresponds to the maximum value of P for a given total energy E and a given total number of particles.

Let us explain this statement by a simple example from games of chance, as is done in probability theory. Let a coin be tossed N times. The probability of its showing heads is 1/2. The probability of its showing heads all N times is $(1/2)^N$. The probability of its showing heads $N - 1$ times and tails once is $N(1/2)^{N-1} \times (1/2)$ since a single showing may occur at any toss, from the first to the last, and the probabilities of alternative events are additive. The probability of getting tails twice is $(1/2)N(N - 1) \times (1/2)^N$.

The first three factors show the number of ways two events can be chosen from a total of N (the number of combinations of N two at

a time). In general, the probability of getting tails k times is

$$q_k = \frac{N!}{k!\,(N-k)!}\left(\frac{1}{2}\right)^{N-k}\left(\frac{1}{2}\right)^k$$

The sum of the all probabilities is

$$\sum_k q_k = \left(\frac{1}{2}\right)^N \left(1 + N + \frac{N\,(N-1)}{1\times 2} + \frac{N\,(N-1)\,(N-2)}{1\times 2\times 3} + \ldots\right)$$

Since the sum of binomial coefficients is 2^N,

$$\sum_k q_k = \left(\frac{1}{2}\right)^N \times 2^N = 1$$

Considering the series q_k for different k's, we see that q_k increases up to the middle, that is, up to $k = N/2$, and then decreases symmetrically with respect to the middle. Indeed, the kth term is obtained from the $(k-1)$st by multiplying by $(N - k + 1)/k$ since the terms increase as long as $N/2 > k$.

Every separate series of tails is in every way equally probable to all the other series. The probability of any given series is $(1/2)^N$. But if we are interested not in the sequence of heads and tails but only in the total number of each, the respective probabilities are q_k. For $N \gg 1$ the function has a sharp maximum at $k = N/2$ and rapidly decreases on both sides of $N/2$. If we call the total number of N tosses a "game", for large N we shall find that tails will occur approximately $N/2$ times in the overwhelming majority of the tossings. The probability maximum is the sharper the greater N is.

Now let us go back to calculating the number of states of an ideal gas.

On the basis of the equiprobability of the direct and reverse transitions between any pair of states we showed that the probabilities of distributions of particles by state for a given total energy is exactly the same. (In the same way, all separate sequences of heads in each separate game are equally probable.) But if we do not specify the state of the gas by indicating which of the g_k states with energy ε_k are occupied and define only the total number of particles in a state with energy ε_k, we obtain a probability distribution with a maximum similar to the probability distribution of games according to the total number of heads irrespective of their sequence. The only difference is that in the game the probability depends on one parameter, k, while the probability of distribution of gas particles by state depends on all n_k's.

Our problem is to find the distribution for particles with integral and half-integral spins. It is more convenient to find the maximum, not of the quantity P itself but of its logarithm; $\ln P$ is a monotonic function and therefore assumes the maximum value at the same time as the argument P does.

Stirling's Formula. In calculations we shall need logarithms of factorials. There is a convenient approximate formula for $\ln n!$, which we shall now develop. It is obvious that

$$\ln n! = \ln (1 \times 2 \times 3 \times 4 \times \ldots \times n) = \sum_{k=1}^{n} \ln k$$

Since the difference $\ln (n+1) - \ln n$ is inversely proportional to n (at $n \gg 1$), logarithms of large numbers vary slowly. Consequently,

$$\ln n! = \sum_{k=1}^{n} \ln k \approx \int_{0}^{n} \ln k \, dk = n \ln n - n = n \ln \frac{n}{e} \tag{1.9}$$

where e is the base of natural logarithms. This is the well-known *Stirling's formula* in a simplified form. The larger the n the better the approximation. Its more precise form is presented in Exercise 1 at the end of this Section.

Accessory Conditions. And so we have to find the numbers n_k for which the quantity

$$S \equiv \ln P = \ln \prod_{k} P(n_k, g_k) \tag{1.10}$$

is a maximum for the given total energy

$$E = \sum_{k} n_k \varepsilon_k \tag{1.11}$$

and the total number of particles

$$N = \sum_{k} n_k \tag{1.12}$$

This kind of an extremum is said to be subject to *constraints* because of the accessory conditions (1.11) and (1.12) imposed on it.

Let us first find n_k for particles not subject to Pauli's exclusion principle, that is, having integral or zero spin. For this first substitute the expression (1.6) into (1.10):

$$S \equiv \ln P = \ln \prod_{k} \frac{(g_k + n_k - 1)!}{n_k! \, (g_k - 1)!} \approx \ln \prod_{k} \frac{(g_k + n_k)!}{n_k! g_k!} \tag{1.13}$$

since the weight of each state, g_k, is considerably greater than unity.

We substitute the factorials in this expression according to Stirling's formula (1.9),

$$S = \sum_{k} \left[(g_k + n_k) \ln \frac{g_k + n_k}{e} - n_k \ln \frac{n_k}{e} - g_k \ln \frac{g_k}{e} \right] \tag{1.14}$$

differentiate (1.14) with respect to all n_k's and equate the differential to zero:

$$dS = \sum_k dn_k \ln \frac{g_k + n_k}{n_k} = 0 \tag{1.15}$$

It cannot be concluded from this equation that the coefficients of all dn_k's are zero, because the n_k's are dependent quantities. The relationships between them are given by (1.11) and (1.12), and in differential form are as follows:

$$dE = \sum_k \varepsilon_k \, dn_k = 0 \tag{1.16}$$

$$dN = \sum_k dn_k = 0 \tag{1.17}$$

These equations could be used to express any two of the numbers dn_k which could be then substituted into (1.15); the remaining n_k's would be independent quantities. The customary procedure, however, is somewhat different.

The Method of Undetermined Multipliers. The elimination of dependent quantities in variational problems is most conveniently accomplished by *Lagrange's method of undetermined multipliers*, which makes it possible to preserve the symmetry among all n_k's. Multiply Eq. (1.16) by an *undetermined multiplier*, which we denote $-1/\theta$. The meaning of this notation will be seen when we compare experimentally observed quantities with the obtained formulas. We multiply Eq. (1.17) by a coefficient which we denote μ/θ. Now combine all three Eqs. (1.15), (1.16), (1.17) and regard all n_k's as independent and θ and μ as unknown quantities that have to be determined from Eqs. (1.11) and (1.12). The maximum condition is now written as

$$dS - \frac{dE}{\theta} + \frac{\mu \, dN}{\theta} = 0 \tag{1.18}$$

We thus look for the extremum of the quantity $S - (E/\theta) + (\mu N/\theta)$ for constant θ and μ, which in turn can be expressed in terms of the total energy and the number of particles. That is the essence of the method of undetermined multipliers in finding an extremum subject to accessory conditions. Having thus got rid of the constraints imposed on the quantities, we can regard them as mutually independent, and any differential dn_k can be assumed zero.

Equation (1.18) written in terms of dn_k has the form

$$dS - \frac{dE}{\theta} + \frac{\mu \, dN}{\theta} = \sum_k dn_k \left(\ln \frac{g_k + n_k}{n_k} + \frac{\mu}{\theta} - \frac{\varepsilon_k}{\theta} \right) = 0 \tag{1.19}$$

The Bose-Einstein Distribution. Assume now that all the differentials except dn_k are equal to zero. From what has just been said this is possible. Then, for Eq. (1.19) to hold, the coefficient of dn_k must be zero:

$$\ln \frac{g_k + n_k}{n_k} + \frac{\mu}{\theta} - \frac{\varepsilon_k}{\theta} = 0 \tag{1.20}$$

The equation, naturally, holds for all k's. Solving it with respect to n_k, we arrive at the most probable distribution of the number of particles of an ideal gas by state:

$$n_k = g_k \left[\exp\left(\frac{\varepsilon_k - \mu}{\theta} \right) - 1 \right]^{-1} \tag{1.21}$$

This formula is called the *Bose-Einstein distribution.* Of the particles for which it holds it is said that they obey *Bose-Einstein statistics* or, for short, *Bose statistics.* They possess integral or zero spin. If they are elementary particles (light quanta, π-mesons, K-mesons, etc.), they are called bosons. The parameters θ and μ introduced into the distribution function can be found in terms of E and N from the equations

$$\sum_k \varepsilon_k g_k \left[\exp\left(\frac{\varepsilon_k - \mu}{\theta} \right) - 1 \right]^{-1} = E \tag{1.22}$$

$$\sum_k g_k \left[\exp\left(\frac{\varepsilon_k - \mu}{\theta} \right) - 1 \right]^{-1} = N \tag{1.23}$$

Thus, the problem of finding the most probable values of n_k is, in principle, solved. It often proves more convenient not to invert Eqs. (1.22) and (1.23), leaving the total energy and number of particles in terms of θ and μ.

The Fermi-Dirac Distribution. Let us now determine the quantity n_k for the case of particles subject to Pauli's exclusion principle. For S we have, from Eq. (1.17) and Stirling's formula,

$$S = \ln \prod_k \frac{g_k!}{n_k!\,(g_k - n_k)!}$$

$$= \sum_k \left[g_k \ln \frac{g_k}{e} - n_k \ln \frac{n_k}{e} - (g_k - n_k) \ln \frac{g_k - n_k}{e} \right] \tag{1.24}$$

Differentiating (1.24) and substituting dS into Eq. (1.18), we obtain

$$dS - \frac{dE}{\theta} + \frac{\mu\, dN}{\theta} = \sum_k dn_k \left(\ln \frac{g_k - n_k}{n_k} - \frac{\varepsilon_k}{\theta} + \frac{\mu}{\theta} \right) = 0 \tag{1.25}$$

whence we arrive at the following extremum condition:

$$\ln \frac{g_k - n_k}{n_k} - \frac{\varepsilon_k}{\theta} + \frac{\mu}{\theta} = 0$$

The required distribution appears thus:

$$n_k = g_k \left[\exp \left(\frac{\varepsilon_k - \mu}{\theta} \right) + 1 \right]^{-1} \tag{1.26}$$

Here, $n_k < g_k$, as it should be with particles subject to Pauli's exclusion principle. Formula (1.26) for such particles is called the *Fermi-Dirac distribution* or, for short, the *Fermi distribution.* The Fermi distribution applies to particles with half-integral spin. If they are elementary particles (for instance, electrons, protons, neutrons), they are called fermions.

The parameters θ and μ are determined from equations analogous to (1.22) and (1.23):

$$\sum_k \varepsilon_k g_k \left[\exp \left(\frac{\varepsilon_k - \mu}{\theta} \right) + 1 \right]^{-1} = E \tag{1.27}$$

$$\sum_k g_k \left[\exp \left(\frac{\varepsilon_k - \mu}{\theta} \right) + 1 \right]^{-1} = N \tag{1.28}$$

The Parameters θ and μ. The parameter θ is an essentially positive quantity, since otherwise it would be impossible to satisfy Eqs. (1.22), (1.23), (1.27), (1.28). Indeed, there is no upper limit to the energy spectrum of gas particles. An infinitely large ε_k and a negative θ would yield $\exp(\varepsilon_k/\theta) = 0$, so that the Bose distribution in itself would lead to the absurd result $n_k < 0$. In (1.23) there would be an infinite quantity, $-\sum g_k$, on the left, which can in no way equal N. Similarly, a Fermi distribution would yield infinite positive quantities $\sum \varepsilon_k g_k$ and $\sum g_k$ in the left-hand sides of (1.27) and (1.28), which is impossible because of the finite values of E and N in the right-hand sides. Therefore

$$\theta \geqslant 0 \tag{1.29}$$

It will be shown in the next section that the quantity θ is proportional to the absolute temperature of the gas.

The parameter μ is of great importance in the theory of chemical and phase equilibria. These applications will be considered later on (see Sec. 8ff.).

The Weight of a State. We present several more formulas for the weight of a state of an ideal gas particle. The weight of a state with an energy lying between ε and $\varepsilon + d\varepsilon$ is given by Eq. [28.25] in which we now denote the left-hand side $dg(\varepsilon)$. Furthermore, as-

sume that the particle possesses an intrinsic angular momentum (spin) **j**, in which case we must take into account the number of possible projections of the vector **j**, which is $2j+1$. Thus

$$dg(\varepsilon)=(2j+1)\frac{Vm^{3/2}\varepsilon^{1/2}\,d\varepsilon}{2^{1/2}\pi^2h^3} \tag{1.30}$$

where V is the volume occupied by the gas. For an electron $j=1/2$ and $2j+1=2$.

For light quanta we must use Eq. [28.24], substituting ω/c for k ($p=hk$) and multiplying by two (the number of possible polarizations of a light quantum for a given wave vector **k**):

$$dg(\omega)=\frac{V\omega^2\,d\omega}{\pi^2c^2} \tag{1.31}$$

It may also be useful to know the weight of a state for which the linear momentum projections are located between p_x and p_x+dp_x, p_y and p_y+dp_y, and p_z and p_z+dp_z. It is determined according to Eq. [28.23], also with account taken of the factor $2j+1$. Thus, for electrons we obtain

$$dg(\mathbf{p})=2\,\frac{V\,dp_x\,dp_y\,dp_z}{(2\pi h)^3} \tag{1.32}$$

EXERCISES

1. Write the formula for the probability that in tossing a coin heads are obtained k times for large N (k is close to the maximum of q_k).

Solution. The general formula is of the form

$$q_k=\frac{N!}{(N-k)!\,k!}\,2^{-N}$$

Assuming N and k large, it is more convenient to apply Stirling's formula in a more exact form than (1.9):

$$\ln N!=N\ln\frac{N}{e}+\frac{1}{2}\ln 2\pi N$$

Assume $k=x+N/2$, $N-k=-x+N/2$, where x is small in comparison with $N/2$. The quantity x can then be neglected in the correction terms $(1/2)\ln 2\pi k$ and $(1/2)\ln 2\pi(N-k)$ in Stirling's formula. Expand the denominator of the expression for q_k in a series up to x^2:

$$\ln(N-k)!=\ln\left(\frac{N}{2}-x\right)!$$

$$=\frac{N}{2}\ln\frac{N}{2e}-x\ln\frac{N}{2}+\frac{x^2}{N}+\frac{1}{2}\ln 2\pi\frac{N}{2}$$

$$\ln k! = \ln\left(\frac{N}{2}+x\right)!$$
$$= \frac{N}{2}\ln\frac{N}{2e} + x\ln\frac{N}{2} + \frac{x^2}{N} + \frac{1}{2}\ln 2\pi\frac{N}{2}$$

The correction term is

$$\frac{1}{2}\left(\ln 2\pi N - 2\ln 2\pi\frac{N}{2}\right) = \frac{1}{2}\ln\frac{2}{\pi N}$$

Substituting into the expression for q_k and taking antilogarithms, we arrive at the required formula:

$$q = \left(\frac{2}{\pi N}\right)^{1/2}\exp\left(-\frac{2x^2}{N}\right)$$

The quantity q has a maximum at $x = 0$ and decreases symmetrically on both sides. It decreases by a factor of e in the interval $x_e = (N/2)^{1/2}$, which characterizes the sharpness of the maximum. The interval x_e constitutes a section of the whole interval of the variation of x, that is, $2x_c/N = (2/N)^{1/2}$. For example, at $N = 1000$ the maximum is approximately 1/40. The ratio x_e/N is about 2% so that in a game of one thousand tosses heads occur basically from 475 to 525 times. The probability of heads (or tails) occurring 400 times out of one thousand is $(1/40)\exp(-2 \times 10\,000/1000) = (1/40)\,e^{-20}$. In other words it is e^{+20} (several hundred million) times less than the probability of its showing 500 times.

2. Show that the sum of all the probabilities of heads occurring computed in the previous problem is, as in the exact formula for q_k, equal to unity, that is, the probability is *normalized* to unity.

Solution. Since the probability decreases very rapidly as the absolute magnitude of x increases, integration can be extended from $-\infty$ to $+\infty$ without appreciable error. As in developing Stirling's formula, the sum can be replaced by an integral

$$\int_{-\infty}^{\infty} q(x)\,dx \left(\frac{2}{\pi N}\right)^{1/2}\int_{-\infty}^{\infty} e^{-2x^2/N}|dx = \frac{1}{\pi^{1/2}}\int_{-\infty}^{\infty} e^{-\xi^2}\,d\xi$$

We calculate the integral

$$I = \int_{-\infty}^{\infty} e^{-\xi^2}\,d\xi$$

Obviously

$$I^2 = \int_{-\infty}^{\infty} e^{-\xi^2}d\xi\int_{-\infty}^{\infty} e^{-\eta^2}\,d\eta = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(\xi^2+\eta^2)}\,d\xi\,d\eta$$

Integration extends over the whole (ξ, η)-surface. Transforming to polar coordinates,

$$\xi = \rho\cos\varphi, \qquad \eta = \rho\sin\varphi$$

we find that $d\xi\, d\eta = \rho\, d\rho\, d\varphi$ and consequently

$$I^2 = \int_0^\infty e^{-\rho^2}\rho\, d\rho \int_0^{2\pi} d\varphi = -2\pi\frac{1}{2}\, e^{-\rho^2}\Big|_0^\infty = \pi$$

or

$$I = \pi^{1/2}$$

Therefore

$$\int_{-\infty}^{\infty} q(x)\, dx = 1$$

3. Find the mean square deviation of the occurrence of heads or tails from the most probable, that is, from $x = 0$.

Solution. We shall denote statistical-average quantities (as distinct from quantum-average) by a short line over the letter that denotes the quantity. The required average of x^2 has the form

$$\overline{x^2} = \int_{-\infty}^{\infty} x^2 q(x)\, dx = \left(\frac{2}{\pi N}\right)^{1/2} \int_{-\infty}^{\infty} x^2 e^{-2x^2/N}\, dx = \frac{N}{2\pi^{1/2}} \int_{-\infty}^{\infty} \xi^2 e^{-\xi^2}\, d\xi$$

To calculate the integral we use the result of the previous problem and write $\alpha\xi^2$ in the exponent instead of ξ^2:

$$\int_{-\infty}^{\infty} e^{-\alpha\xi^2}\, d\xi = \left(\frac{\pi}{\alpha}\right)^{1/2}$$

Differentiation of both sides with respect to α yields

$$-\int_{-\infty}^{\infty} \xi^2 e^{-\alpha\xi^2} d\xi = -\frac{\pi^{1/2}}{2\alpha^3}$$

Assuming $\alpha = 1$, we obtain

$$\int_{-\infty}^{\infty} \xi^2 e^{-\xi^2}\, d\xi = \frac{\pi^{1/2}}{2}$$

We note incidentally that

$$\int_{-\infty}^{\infty} \xi^4 e^{-\xi^2}\, d\xi = \frac{3\pi^{1/2}}{4}$$

and in general

$$\int_{-\infty}^{\infty} \xi^{2n} e^{-\xi^2}\, d\xi = \frac{1\times 3\times 5\times \ldots \times (2n-1)}{2n}\, \pi^{1/2}$$

Taking into account the obtained integral, we have

$$\overline{x^2} = N/4$$

Expressing N in terms of $\overline{x^2}$, we can write the probability distribution law as follows:

$$q(x)\,dx = \frac{1}{(2\pi\overline{x^2})^{1/2}}\, e^{-x^2/2\overline{x^2}}$$

Consequently, the distribution width x_e is simply related to the mean square deviation of x from its most probable value:

$$x_e = (2\overline{x^2})^{1/2}$$

At $x = x_e$ the probability $q(x)$ decreases e times compared with $q(0)$. Of course, the obtained relationship between x_e and $\overline{x^2}$ holds for the exponential distribution obtained here and for certain other specially selected distributions, but not in the most general case. The expression for $q(x)$ is called the *Gaussian distribution.*

2

BOLTZMANN STATISTICS: TRANSLATIONAL MOTION OF MOLECULES; GAS IN AN EXTERNAL FIELD

The Boltzmann Distribution. Long before the Bose and Fermi quantum distribution formulas (1.21) and (1.26) were obtained, Ludwig Boltzmann enunciated the classical energy distribution law for the molecules of an ideal gas. This law is obtained from both quantum distributions by means of a limiting process. We shall first carry out the transition in a purely formal way and then examine the real conditions it corresponds to.

Let ε be measured from zero and let the ratio μ/θ be negative and large in absolute value. Then

$$\exp\left(-\frac{\mu}{\theta}+\frac{\varepsilon}{\theta}\right)$$

is much larger than unity for all ε. In comparison with this quantity unity can be neglected. In that case the Bose and Fermi distributions assume the same form

$$n_k = g_k \exp\left(\frac{\mu-\varepsilon_k}{\theta}\right) \tag{2.1}$$

This is the *Boltzmann distribution.* Let us now determine the constant μ from the conditions (1.23) or (1.28), which in the limit

reduce to the same formula

$$\sum_k n_k = \sum_k g_k \exp\left(\frac{\mu - \varepsilon_k}{\theta}\right) = N \tag{2.2}$$

Let us assume that in addition to the external translational degrees of freedom the gas molecules possess some internal degrees of freedom. They may be related to electron excitation, the vibration of nuclei with respect to one another, or the rotation of the molecule as a whole in space. The energy of all these degrees of freedom is quantized. Without defining it more precisely for the time being, we can write the total energy ε of a molecule as the sum of the energies of translational and internal motion:

$$\varepsilon = \frac{p^2}{2m} + \varepsilon^{(i)} \tag{2.3}$$

Correspondingly, the weight of the state with given energy can be represented as the product of two weights: one related to the translational motion and given by Eq. (1.32); the other we denote simply $g^{(i)}$, agreeing to include in it the factor $2j + 1$. Thus

$$dg(\mathbf{p}) = \frac{V\, dp_x\, dp_y\, dp_z}{(2\pi h)^3} g^{(i)} \tag{2.4}$$

Therefore Eq. (2.2) can be written thus:

$$\frac{V}{2\pi h^3} e^{\mu/\theta} \sum_i g^{(i)} e^{\ \varepsilon^{(i)}/\theta} \int\int\int_{-\infty}^{\infty} \exp\left(-\frac{p^2}{2m\theta}\right) dp_x\, dp_y\, dp_z = N \tag{2.5}$$

Representing the energy of translational motion as $(1/2m)(p_x^2 + p_y^2 + p_z^2)$ and taking into account that integration of each of the three components of linear momenta from $-\infty$ to $+\infty$ is carried out independently, we note that the triple integral in (2.5) can be represented as the product of three integrals of the form

$$\int_{-\infty}^{\infty} \exp\left(-\frac{p_x^2}{2m\theta}\right) dp_x$$

The method of calculating such an integral was shown in the preceding section (see Exercise 3 of the preceding section). We obtain for it the value $(2\pi m\theta)^{1/2}$. Consequently, the condition (2.5) reduces to the form

$$e^{-\mu/\theta} = \frac{V}{N}\left(\frac{m\theta}{2\pi}\right)^{3/2} \frac{1}{h^3} \sum_i g^{(i)} e^{-\varepsilon^{(i)}/\theta} \tag{2.6}$$

If the gas is monatomic, the quantities $\varepsilon^{(i)}$ refer to electronic excitations. If $\varepsilon^{(i)} \gg \theta$, virtually only the zero term appears in the

summation over the states.[3] But since the energy is measured from $\varepsilon^{(0)}$ as from zero, the whole summation reduces practically to the one term $g^{(0)}$ whose value is of the order of unity. For example, when the ground state has angular momentum 1/2, $g^{(0)} = 2$, and the condition for the applicability of the Boltzmann statistics takes the form:

$$-\frac{\mu}{\theta} = \ln\left[\frac{g^{(0)}}{h^3}\frac{V}{N}\left(\frac{m\theta}{2\pi}\right)^{3/2}\right] \gg 1 \tag{2.7}$$

For the inequality (2.7) to be satisfied it is sufficient that one of two conditions hold: (1) the density of the gas is very small, that is, the volume occupied by the gas at the given temperature θ is large; (2) the temperature θ for a given volume V is very high.

When the gas is not monatomic, these conditions change somewhat in quantitative terms, since the sum over discrete states in Eq. (2.6) is also dependent on θ. Qualitatively, however, the conditions of applicability of the Boltzmann statistics hold.

Classical and Quantum Statistics. We have found that at low densities and high temperatures the quantum distribution laws for gases pass into the classical Boltzmann law. We shall, from now on, agree to call the Bose and Fermi statistics *quantum statistics*, and the Boltzmann statistics, *classical*, regardless of whether the energy spectrum is continuous or discrete. Quantum statistics are those that take account of the indistinguishability of identical particles. In other words, quantum statistics is based on the quantum definition of the state of a system: the number of particles in all quantum states must be given.

The classical definition of the state of a system indicates which particles are in the given states, since it is possible (in principle) to trace their paths. The Boltzmann formula can be derived from the classical definition directly, bypassing quantum laws.

The Maxwell Distribution. In this section we shall not concern ourselves with the statistics of internal motion of molecules. Equation (2.1) here applies only to their translational motion in space. From Eq. (2.3) the energy of translational motion is separable from their internal energy. Therefore the Boltzmann distribution separates into the product of two factors. The factor relating to translational motion has the form

$$\exp\left(-\frac{p^2}{2m\theta}\right)$$

---

[3] The relation between θ and temperature is given by Eq. (2.25).

The weight of a state relating to a given absolute value is obtained by transition to polar coordinates in Eq. (1.32), which yields

$$dg(p) = \frac{Vp^2\,dp}{(2\pi h)^2} \tag{2.8}$$

(see [28.24]).

Thus, the distribution of the absolute value of the linear momentum is written in the form:

$$dn(p) = A\exp\left(-\frac{p^2}{2m\theta}\right)p^2\,dp \tag{2.9}$$

It is applicable to both monatomic and polyatomic gases if m is taken as the mass of the molecule as a whole.

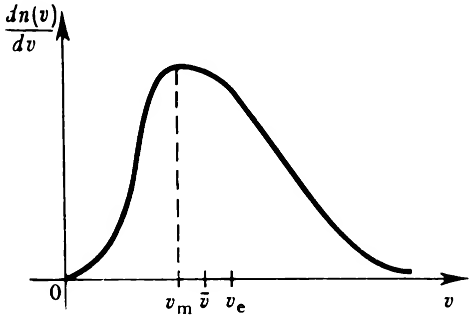


Figure 1

The constant A is determined by the condition (2.2), that is, by normalizing the distribution over the total number of particles:

$$A\int_0^\infty p^2\exp\left(-\frac{p^2}{2m\theta}\right)dp = N \tag{2.10}$$

The value of the integral was calculated in Exercise 3 of the preceding section. We thus find that

$$A = N\left(\frac{2}{\pi}\right)^{1/2}\frac{1}{(m\theta)^{3/2}} \tag{2.11}$$

In place of the momentum distribution of molecules it is sometimes useful to have their velocity distribution. For this it is sufficient to substitute $p = mv$ into the distribution (2.9):

$$dn(v) = N\left(\frac{2}{\pi}\right)^{1/2}\left(\frac{m}{\theta}\right)^{3/2}\exp\left(-\frac{mv^2}{2\theta}\right)v^2\,dv \tag{2.12}$$

This distribution was developed by James Clerk Maxwell, before Boltzmann, which is why it is called the *Maxwell distribution* (see Figure 1).

In Figure 1 the derivative $dn(v)/dv$ has been plotted on the vertical axis. For small values of v this quantity is close to zero because of the factor v^2 in the formula for the weight of a state; then it attains a maximum, from which it tends exponentially to zero at high velocities. Thus, a gas contains molecules with all velocity values.

The greatest number of molecules have velocities corresponding to the maximum of the distribution curve. This maximum is given by Eq. (2.12). The corresponding velocity is termed the *modal* (most probable) *velocity* and is equal to

$$v_m = \left(\frac{2\theta}{m}\right)^{1/2} \tag{2.13}$$

The *mean velocity* is

$$\begin{aligned}\bar{v} &= \left(\frac{2m^3}{\pi\theta^3}\right)^{1/2} \int_0^\infty \exp\left(-\frac{mv^2}{2\theta}\right) v^3\, dv \\ &= \frac{1}{2}\left(\frac{2m^3}{\pi\theta^3}\right)^{1/2} \left(\frac{2\theta}{m}\right)^2 = \left(\frac{8\theta}{\pi m}\right)^{1/2}\end{aligned} \tag{2.14}$$

The *mean square velocity* is

$$\begin{aligned}\overline{v^2} &= \left(\frac{2m^3}{\pi\theta^3}\right)^{1/2} \int_0^\infty \exp\left(-\frac{mv^2}{2\theta}\right) v^4\, dv \\ &= \left(\frac{2m^3}{\pi\theta^3}\right)^{1/2} \left(\frac{2\theta}{m}\right)^{5/2} \frac{3\pi^{1/2}}{8} = \frac{3\theta}{m}\end{aligned} \tag{2.15}$$

and $(\overline{v^2})^{1/2} = v_e$ (the *effective velocity*). Here we make use of the result obtained in Exercise 3 of the preceding section. Note that $v_e : \bar{v} : v_m = \sqrt{3} : \sqrt{8/\pi} : \sqrt{2}$.

The *mean energy per molecule* is

$$\bar{\varepsilon} = \frac{1}{2} m\overline{v^2} = \frac{3}{2}\theta \tag{2.16}$$

and the *mean energy of the gas as a whole* is N times greater:

$$\bar{E} = \frac{3}{2} N\theta \tag{2.17}$$

This is the energy of translational motion of the molecules. Now we shall determine the numerical values expressed by Eqs. (2.13)-(2.15).

The Relationship Between Energy Density and Pressure. We shall now derive a very important relationship between the energy density of translational motion of the molecules of a gas and its pressure, which holds for all statistics and depends only on the form of the expression for energy in terms of momentum.

The pressure of a gas is defined as the force with which the gas acts upon a unit area perpendicular to the direction of the force. This force, in turn, is equal to the component of the momentum normal to the surface and transmitted by the gas molecules in unit time. Let the direction of the normal to the surface coincide with the x axis. First we choose those molecules which have a velocity component v_x along the x axis. If initially they were located in a layer of width v_x, they will reach the surface in unit time (if, for example, $v_x = 100$ m-s^{-1}, then obviously in one second the molecules from a layer 100 metres thick will reach the surface). Now out of this layer we cut a cylinder with a base of unit area and a height equal to v_x. The volume of such a cylinder is v_x. If $dn(v_x)$ is the number of molecules whose velocity component normal to the surface is v_x, the density of such molecules is $(1/V)\,dn(v_x)$. In a cylinder of volume v_x there are $(v_x/V)\,dn(v_x)$ such molecules. Each of them after elastically colliding with the base will reverse its normal velocity component, and the base will receive a momentum

$$mv_x - (-mv_x) = 2mv_x \tag{2.18}$$

Thus all the gas molecules having a velocity v_x will in unit time transfer to the base the momentum

$$2mv_x \frac{dn\,(v_x)}{V} v_x = 2mv_x^2 \frac{dn\,(v_x)}{V} \tag{2.19}$$

In order to obtain the total pressure of the gas on the base we must integrate (2.19) over v_x from 0 to ∞ (but not from $-\infty$ to $+\infty$, since the molecules moving away from the base do not strike it). Thus, the pressure of the gas on the wall is

$$p = \frac{2m}{V} \int\limits_0^\infty v_x^2\, dn\,(v_x) \equiv \frac{m}{V} \int\limits_{-\infty}^\infty v_x^2\, dn\,(v_x) \tag{2.20}$$

On the other hand, the mean kinetic energy of the gas is

$$\overline{E} = \frac{m}{2} \left[\int\limits_{-\infty}^\infty v_x^2\, dn\,(v_x) + \int\limits_{-\infty}^\infty v_y^2\, dn\,(v_y) + \int\limits_{-\infty}^\infty v_z^2\, dn\,(v_z) \right]$$

$$= \frac{3m}{2} \int\limits_{-\infty}^\infty v_x^2\, dn\,(v_x) \tag{2.21}$$

(The mean squares of all the velocity components are equal.)

Comparing (2.20) with (2.21), we find that the pressure of the gas equals 2/3 the density of its mean kinetic energy:

$$p = \frac{2}{3} \frac{\overline{E}}{V} \tag{2.22}$$

This result was published by Daniel Bernoulli in 1738, one hundred and fifty years before statistical physics became an independent science.

Only two assumptions were used in deriving Eq. (2.22): (1) equal values of the three velocity projections are equiprobable and (2) if the momentum is mv, the kinetic energy is $mv^2/2$. The concrete form of the distribution function is immaterial.

If a gas obeys the Boltzmann statistics, the mean kinetic energy of the gas, $\bar{E}$, is equal to $3N\theta/2$, according to (2.17). Substituting this into Eq. (2.22), we obtain

$$pV = N\theta \tag{2.23}$$

On the other hand, we have the following definition of absolute temperature from the *ideal gas law*

$$pV = RT \tag{2.24}$$

The quantities on the left-hand side of this equation have no relationship to heat measurements, therefore an ideal gas can be used as a thermometric substance. If the constant R for one mole of gas is taken equal to 8.314×10^7, the temperature T is expressed in *kelvins* (K).

On the other hand, comparing Eqs. (2.23) and (2.24), we find the relationship between the "statistical" temperature θ involved in the distribution function and having the dimension of energy (erg) and the temperature T according to the Kelvin scale:

$$\theta = \frac{RT}{N_A} = \frac{8.314 \times 10^7 \text{ erg-K}^{-1}\text{mol}^{-1}}{6.023 \times 10^{23} \text{ mol}^{-1}} T = k_B T \tag{2.25}$$

where N_A is *Avogadro's number*.

The ratio $k_B \equiv R/N_A$ is called *Boltzmann's constant*. It is equal to 1.38×10^{-16} erg-K^{-1}. Temperature can also be measured in electron volts, one electron volt being equal to 1.60×10^{-12} erg. Translating ergs into kelvins with the help of Boltzmann's constant, we find that 1 eV = 11 600 K.

The derivative of energy with respect to temperature for constant volume is called the *specific heat* c_V. For an ideal monatomic gas it equals $3R/2$, which corresponds to an energy $3RT/2$. Replacing RT by $N_A\,\theta$, we obtain $\bar{E} = 3N_A\theta/2$ in agreement with Eq. (2.17).

The relationship (2.25) allows us to calculate the mean velocities of molecules without using Avogadro's number:

$$\bar{v} = \left(\frac{8\theta}{\pi m}\right)^{1/2} = \left(\frac{8RT}{\pi N_A m}\right)^{1/2} = \left(\frac{8RT}{\pi M}\right)^{1/2}$$

where M is the molecular weight of the gas. For example, the mean velocity of a hydrogen molecule at 300 K is

$$\bar{v} = \left(\frac{8 \times 8.3 \times 10^7 \times 300}{3.14 \times 2} \right)^{1/2} \approx 1800 \text{ m-s}^{-1}$$

This value is comparable with the velocity of sound (see Sec. 16).

Thermonuclear Reactions. When nuclei collide, reactions are possible between them that proceed with the release of energy. For example, in a deuteron-deuteron collision one of two reactions can occur (besides elastic scattering):

$${}_1D^2 + {}_1D^2 \rightarrow \begin{cases} {}_2He^3 + {}_0n^1 \\ {}_1H^3 + {}_1H^1 \end{cases}$$

Here ${}_1H^3$ is tritium, ${}_2He^3$ is the light isotope of helium, ${}_0n^1$ is the neutron, ${}_1H^1$ is the proton, and ${}_1D^2 \equiv {}_1H^2$ is the deuteron. Another example is

$${}_3Li^6 + {}_1D^2 \rightarrow 2{}_2He^4$$

For charged nuclei to collide effectively they must overcome the potential barrier of Coulomb repulsion, which was examined in [Sec. 31]. The probability of penetrating the potential barrier depending on energy is basically determined by the barrier factor

$$\exp\left(-\frac{2\pi Z_1 Z_2 e^2}{hv} \right) \tag{2.26}$$

(see the first term on the right in [31.27]).

Here, Z_1e and Z_2e are the charges of the colliding nuclei, and v is their relative velocity (displacement of their common centre of mass does not result in collisions and, consequently, reactions).

A reaction can be produced by accelerating particles in a discharge tube. But when charged particles enter a substance, they dissipate their energy mainly on exciting and ionizing the atoms of the substance. Not more than one out of 10^5 or 10^6 of the bombarding particles trigger a reaction. Therefore the energy yield of the reaction is substantially less than the total energy expended on accelerating the beam of particles.

Things are different if a substance capable of entering a reaction is heated to a very high temperature of the order of 10^7 K (10^3 eV). At this temperature nuclei already react at a considerable rate (transfer of energy to electrons does not occur since they are separated from the nuclei by thermal ionization and possess the same mean energy as the nuclei).

Let us calculate the rate of a nuclear reaction occurring under such conditions (such a reaction is called *thermonuclear*). Let the effective cross section of the reaction between nuclei with relative

velocity v be $\sigma(v)$, and let us assume that the nuclei are different; we shall call them *1* and *2*. Construct on every nucleus *2* a cylinder with a base area $\sigma(v)$ and a height numerically equal to v. Then, by difinition of $\sigma(v)$, in unit time all nuclei *1* located in the volume of these cylinders have the velocity v with respect to nuclei *2* and will be involved in the reaction (see [Sec. 6]).

The number of such events per unit volume per unit time is equal to the product

$$n_1 v\sigma(v) \times n_2\, dq(v) \tag{2.27}$$

where n_1 and n_2 are the numbers of nuclei *1* and *2* in unit volume, and $dq(v)$ is the probability that the relative velocity is v. If *1* and *2* are nuclei of the same type (identical), the expression (2.27) should be halved so as not to count each reaction twice. We denote this by the factor 2 in the denominator of Eq. (2.28).

Now let us determine the probability factor $dq\,(v)$. The absolute velocity distribution is given by the product of Maxwell factors of the form

$$\exp\left(-\frac{m_1v_1^2}{2\theta}\right)\times\exp\left(-\frac{m_2v_2^2}{2\theta}\right)=\exp\left[-\frac{1}{2\theta}\left(m_1v_1^2+m_2v_2^2\right)\right]$$

In the exponent of this expression is the sum of the kinetic energies of both nuclei. According to Eq. [3.17], it can be separated into the kinetic energy of motion of the centre of mass of the nuclei and the kinetic energy of their relative motion. Hence, we can separate the factor which yields the relative-velocity distribution,

$$\exp\left(-\frac{m}{2\theta}v_0^2\right)$$

where m is the reduced mass of the nuclei equal to $m_1m_2/(m_1+m_2)$ [3.20]. The relative velocity v_0 is equal to $|\,\mathbf{v}_1-\mathbf{v}_2\,|$. For collisions only the velocity component along the line connecting the two nuclei is important. If we resolve it into two mutually perpendicular components $\mathbf{v}'$ and $\mathbf{v}$,

$$\mathbf{v}_0=\mathbf{v}+\mathbf{v}'$$

the volume element in the velocity space of $\mathbf{v}_0$ will be the product $2\pi v'\,dv'\,dv$. Separating the factors of the distribution function and the volume element depending on v', we are left with only the distribution $dq(v)$ necessary to calculate the rate of the reaction. Normalized to unity the distribution has the form

$$dq(v)=\left(\frac{2m}{\pi\theta}\right)^{1/2}\exp\left(-\frac{mv^2}{2\theta}\right)dv\quad(0\leqslant v\leqslant\infty)$$

The barrier factor (2.26) is also dependent on v.

Thus, the rate of a thermonuclear reaction is

$$k = \frac{n_1 n_2}{2} \int_0^\infty \sigma(v)\, v\, dq(v)\ \frac{\text{events}}{\text{cm}^3\text{s}} \tag{2.28}$$

In this expression the effective cross section barrier factor is

$$\sigma(v) = \sigma_0(v) \exp\left(-\frac{2\pi Z_1 Z_2 e^2}{hv}\right)$$

The factor $\sigma_0(v)$ depends on the rate much less than the barrier factor.

The integral in (2.28) reduces to the form

$$\int_0^\infty \sigma_0(v)\, v \exp\left(-\frac{2\pi Z_1 Z_2 e^2}{hv} - \frac{mv^2}{2\theta}\right) dv \tag{2.29}$$

It can be calculated to a good approximation when the temperature is so low that only the fastest nuclei are capable of reacting. They correspond to the "tail" of the Maxwell distribution in Figure 1. At higher temperatures, when the barrier factor takes on a value of the order of unity, it is of no great significance already at the most probable velocity. Let us now examine the approximate method of calculating the integral (2.29).

We denote the exponent under the integral as

$$f(v) = \frac{2\pi Z_1 Z_2 e^2}{hv} + \frac{mv^2}{2\theta} = \frac{a}{v} + \frac{bv^2}{2}$$

where $a = 2\pi Z_1 Z_2 e^2/h$ and $b = m/\theta$. We find the minimum of the function $f(v)$ from the condition

$$\frac{df}{dv} = -\frac{a}{v^2} + bv = 0, \qquad v_{\min} = \left(\frac{a}{b}\right)^{1/3} \tag{2.30}$$

Let us show that the main contribution to the integral is given by values of v close to $v_{\min}$. Near the minimum the function $f(v)$ can be represented in the form

$$\begin{aligned} f(v) &= f(v_{\min}) + \frac{1}{2}(v - v_{\min})^2 \left(\frac{d^2 f}{dv^2}\right)_{v=v_{\min}} \\ &= \frac{3}{2}(a^2 b)^{1/3} + \frac{3}{2} b (v - v_{\min})^2 \end{aligned} \tag{2.31}$$

Accordingly, the integral (2.29) takes the form

$$\int_0^\infty \sigma_0(v)\, v \exp\left[-f(v_{\min}) - \frac{3}{2} b (v - v_{\min})^2\right] dv \tag{2.32}$$

Since $f(v)$ enters the exponent with a minus sign, the minimum of $f(v)$ corresponds to the velocity $v_{\min}$ of the nuclei at which the greatest number of reactions occur.

Taking into account that

$$\bar{v} = \left(\frac{2m}{\pi\theta}\right)^{1/2} \int_0^\infty v \exp\left(-\frac{mv^2}{2\theta}\right) dv = \left(\frac{2\theta}{\pi m}\right)^{1/2} = \left(\frac{2}{\pi b}\right)^{1/2}$$

the ratio of the velocity $v_{\min}$ to the mean relative velocity $\bar{v}$ can be represented in the form

$$\frac{v_{\min}}{\bar{v}} = \left(\frac{\pi}{2}\right)^{1/2} (a^2 b)^{1/6} = \left(\frac{\pi}{2}\right)^{1/2} \left(\frac{2\pi Z_1 Z_2 e^2 m^{1/2}}{h\theta^{1/2}}\right) \qquad (2.33)$$

We shall call the temperature low if the ratio $v_{\min}/\bar{v}$ is several times greater than unity. At low temperatures the maximum of the integrand (2.32) is very sharp. Indeed, the value of the integrand decreases e times when v deviates from $v_{\min}$ by $(2/3b)^{1/2}$, which by definition is considerably less than $v_{\min} = (a/b)^{1/3}$.

Consequently, we can justifiably cut off the expansion (2.31) at the second term. Besides, the quantities $\sigma_0(v)$ and v can be taken outside the integral sign at $v = v_{\min}$. The error of both approximations is of the order $\bar{v}/v_{\min}$. The integration can be taken from $-\infty$ to $+\infty$ (the integrand decreases rapidly as we recede from $v_{\min}$):

$$\int_0^\infty \sigma_0(v)\, v \exp\left[-f(v_{\min}) - \frac{3}{2} b (v - v_{\min})^2\right] dv$$

$$\approx \sigma_0(v_{\min})\, v_{\min} \exp[-f(v_{\min})] \int_{-\infty}^{\infty} \exp\left[-\frac{3}{2} b (v - v_{\min})^2\right] dv$$

$$= \sigma_0(v_{\min}) \left(\frac{2\pi}{3b}\right)^{1/2} \exp[-f(v_{\min})] \qquad (2.34)$$

Substituting the values of a and b and referring to (2.28), we find the expression for the rate of a thermonuclear reaction:

$$k = \frac{n_1 n_2}{2 \times 3^{1/2}} v_{\min} \sigma_0(v_{\min}) \exp\left[-\frac{3}{2}\left(\frac{m}{\theta}\right)^{1/3} \left(\frac{2\pi Z_1 Z_2 e^2}{h}\right)^{2/3}\right]$$

$$v_{\min} = \left(\frac{2\pi Z_1 Z_2 e^2 \theta}{hm}\right)^{1/3} \qquad (2.35)$$

The exponential factor depends very greatly on temperature. For example, for a thermonuclear reaction in deuterium this factor changes by 3600 times when the temperature increases from 100 to 200 eV. The obtained formula corresponds to the conditions of *slow thermonuclear reactions*. These conditions occur naturally within stars.

Ideal Gas in an External Potential Field. We shall now consider an ideal gas in an external field with a potential U. Potential energy can depend on the location of a molecule's centre of mass in space, its orientation with respect to the external field (if the gas is not monatomic), and the projection of the molecule's spin on the field.

The total energy of a molecule is

$$\varepsilon = \varepsilon^{(i)} + \frac{p^2}{2m} + U \tag{2.36}$$

If U depends on the molecule's position is space, that is $U = U(x, y, z)$, we must transform from the finite volume V in (2.4) to an infinitely small volume $dV = dx\,dy\,dz$. Then the part of the distribution function that depends on the coordinates can be separated, yielding a formula describing the dependence of gas density on position in space:

$$dn(x, y, z) = n_0 \exp\left[-U(x, y, z)/\theta\right] dx\,dy\,dz \tag{2.37}$$

Here the potential is subject to the accessory condition $U(0, 0, 0) = 0$, and n_0 is the gas density at point (0, 0, 0). Obviously, in a gravitational field for which $U = mgz$, we obtain

$$dn(z) = n_0 \exp(-mgz/\theta)\,dz \tag{2.38}$$

It should be noted that in the earth's atmosphere the *barometric height formula* (2.38) is not exact since the temperature of the air varies with height.

In addition, the barometric formula indicates that the air composition must vary with height owing to the different molecular weights of the atmospheric gases (nitrogen, oxygen, etc.). Actually, the altitude composition of the atmosphere is almost uniform owing to intensive mixing processes.

The Nonequilibrium State of Planetary Atmospheres. Let us substitute the exact expression [3.5] for the approximate expression for potential energy in a gravitational field. First express the constant a in [3.5] in terms of more convenient quantities. The force of gravity at the surface of the earth is equal to $-mg$; from the law of universal gravitation it is also equal to $-a/(r_0)^2$, where r_0 is the radius of the earth. Hence, $a = mg(r_0)^2$ and $U = -mg(r_0)^2/r$.

The gas density must then change with altitude according to the law

$$n = n_\infty \exp\left[-mg\,(r_0)^2/(r\theta)\right] \tag{2.39}$$

This function remains finite even at infinite distances from the earth; and since the exponent at infinity is equal to zero, in accordance with the boundary condition $U(\infty) = 0$, the proportionality factor is n_∞.

Near the earth, where $r = r_0$, the density is greater than at infinity by as many times as

$$\exp \frac{mgr_0}{\theta} = \exp \frac{Mgr_0}{RT}$$

is greater than unity.

Since $r_0 \approx 6.4 \times 10^8$ cm and $g = 10^3$ cm-s^{-2}, for oxygen

$$\frac{Mgr_0}{RT} \approx \frac{32 \times 10^3 \times 6.4 \times 10^8}{8.3 \times 10^7 \times 300} \approx 800$$

Obviously, at infinity the density of the terrestrial atmosphere is zero. It therefore follows from Eq. (2.39) that in the earth's gravitational field the atmosphere cannot arrive at the most probable state and gradually dissipates in space. However, at infinity the density of the atmosphere in its most probable state would have to be e^{800} times less than the density at the surface of the earth, and therefore the present state of the atmosphere is very close to the most probable. For the moon that state has been reached: its atmosphere has dissipated completely (if it ever existed).

There is another simple way of explaining the cause of the dispersion of gas. Every particle whose velocity exceeds 11.2 km-s^{-1} is capable of overcoming the earth's gravity (the *escape velocity*). The motion of such a particle is infinite. According to the Maxwell distribution (4.12), in a gas there are always molecules with any velocity. In literal notation, the velocity of molecules capable of escaping into infinity is given by the relationship

$$\frac{1}{2} mv^2 \geqslant mgr_0 \tag{2.40}$$

(the kinetic energy of a molecule at the earth's surface is equal to or greater than the potential energy taken with the opposite sign). Substituting the smallest of these velocities into the Maxwell distribution, we obtain $\exp(-mgr_0/\theta)$ for the fraction of molecules capable of leaving the atmosphere. It is easy to estimate the number of such molecules in the atmosphere at any given time. The area of the earth's surface is 5×10^{18} cm^2. There is 1030 grams of air, or 35 moles, over every square centimetre. Hence, the total number of molecules in the atmosphere is $5 \times 10^{18} \times 35 \times 6 \times 10^{23} = 10^{44}$, and the fraction of molecules possessing velocities exceeding 11.2 km-s^{-1} is $e^{-800} = 10^{-344}$. Therefore the average number of molecules capable of escaping the earth is only 10^{-300}. The proportion for hydrogen ($M = 2$) is quite different. Instead of the exponent —344, the exponent is equal to —21. It is not surprising then that there is practically no hydrogen in the atmosphere.

It should also be noted that molecules close to the surface of the earth cannot carry their energy to the upper atmosphere owing to collisions with other molecules.

EXERCISES

1. Find the mean relative velocities of two molecules in a gas **mixture.**

Solution. The relative-velocity distribution is given by a formula similar to the v-distribution in the problem on the thermonuclear reaction. Instead of the mass of one molecule we must introduce into the distribution the reduced mass of two molecules, $m_1m_2/(m_1+m_2)$. Then the mean relative velocity is, from (2.14),

$$v_0=\left(\frac{8\theta}{\pi m}\right)^{1/2}=\left(\frac{8\theta\,(m_1+m_2)}{\pi m_1 m_2}\right)^{1/2}$$

If the molecules are identical, their mean relative velocity is $2^{1/2}$ times greater than the mean absolute velocity.

2. Calculate the rate k' of a bimolecular reaction if the effective cross section depends on the relative-velocity component along the line joining the molecules in the following way:

$$\begin{aligned}\sigma(v)&=0 \quad \text{if} \quad v<(2A/m)^{1/2}\\ &=\sigma_0 \quad \text{if} \quad v\geqslant(2A/m)^{1/2}\end{aligned}$$

Solution. From the general formula (2.28) we find that

$$k'=\frac{n_1n_2}{2}\left(\frac{m}{2\pi\theta}\right)^{1/2}\int\limits_{(2A/m)^{1/2}}^{\infty}\sigma_0 v\exp\left(-\frac{mv^2}{2\theta}\right)dv$$

$$=\frac{n_1n_2}{2}\left(\frac{\theta}{2\pi m}\right)^{1/2}\sigma_0 e^{-A/\theta}$$

(the *Arrhenius equation*).

The decisive quantity in this result is the exponential factor $\exp(-A/\theta)$. Quantity A is called the *activation energy*. It is equal to the height of the potential barrier over which the colliding particles must pass for the reaction to occur. It is assumed here that the reacting particles obey the laws of classical mechanics. In this problem, transitions below the barrier make a vanishingly small contribution. The obtained formula holds only for exchange reactions of the type $AB+CD=AC+BD$, in which the reaction products carry over the surplus energy yielded in the reaction (not to be confused with the activation energy) if the reaction is exothermic. For an exothermic reaction of the type $A+B=AB$ to occur the liberated energy must be carried over by a third particle involved in the collision but not in the reaction. The pre-exponential factor for a triple collision is not of the type of the factor in the Arrhenius equation.

3

BOLTZMANN STATISTICS: VIBRATIONAL AND ROTATIONAL MOLECULAR MOTION

Molecular Energy Levels. In order to apply statistics to gases consisting of molecules we must classify the energy levels of the molecules. The fact that nuclei are much heavier than electrons and

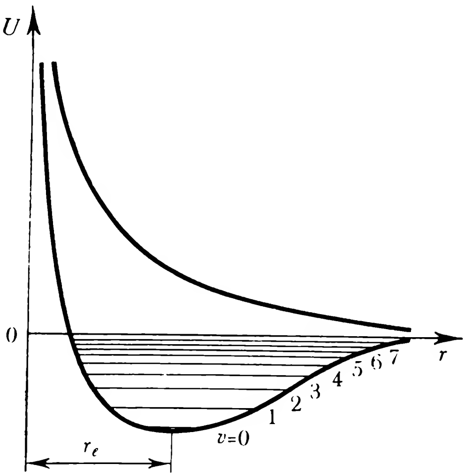


Figure 2

therefore travel much slower is of great help in solving this problem. We used this in [Sec. 34] when considering the binding energy of two hydrogen atoms in a hydrogen molecule. In a diatomic molecule the position of the nuclei is determined by a single parameter, the distance between them, on which the energy eigenvalue of the electrons depends. Addition of the Coulomb repulsion energy of the nuclei and the rotational energy of the electrons in space to the electron energy yields, for a given electron wave function, the energy of the molecule as a function of the distance between the nuclei. For example, in a hydrogen molecule the curves representing this relationship are of different form for parallel and antiparallel electron spin orientations (Figure 2). The lower curve refers to the state with a symmetrical spatial wave function and antiparallel spins, and the upper curve refers to the state with an antisymmetrical spatial

function and parallel spins. The lower curve has a minimum at $r = r_e$, that is, hydrogen atoms may form a molecule only in a definite electron state.

In the general case the potential curves of different electron states can have a minimum. The distances between the curves are given by wave equations of the type [34.7]. In this equation we can neglect the terms containing the masses of the nuclei in the denominators. Hence, the energy gap between different electronic states of the molecules is the same as for an atom, that is, from one to ten electron volts.

Close to the minimum of potential energy the nuclei may perform small oscillations. To the first approximation, these oscillations are harmonic, so that their energy is given by the general equation

$$\varepsilon_{\text{vib}} = h\omega_{\text{vib}}\left(v + \frac{1}{2}\right) \tag{3.1}$$

(see [27.23] and [27.28*b*]).

Here v is called the *vibrational quantum number* of the molecule. It is naturally an integer. Figure 2 shows a more general dependence of energy on r, taking into account that the potential energy curve is not a parabola. The energy levels for such cases were found in [Sec. 29]. In practice deviations from Eq. (3.1) have little effect on statistical quantities, because when oscillations with large values of v are excited, dissociation occurs.

The frequency ω_{vib} depends on the electron state in which nuclear oscillations occur. In accordance with the general formula [7.12], we obtain for frequency the expression

$$\omega_{\text{vib}} = \left[\frac{1}{m}\left(\frac{d^2U}{dr^2}\right)_{r=r_e}\right]^{1/2}$$

It will be observed that the frequency is inversely proportional to the square root of the reduced mass of the nuclei. Therefore the vibrational energy quantum is considerably less than the distance between electron levels, which are independent of the nuclear mass. The quantity $h\omega$ is of the order of tenths of an electron volt.

In addition to vibrational motion a diatomic molecule may also perform rotational motion as a whole. Rotation is most simply taken into account when the total spin of the electrons is zero. In the ground state of the molecule the projection of the total orbital angular momentum of the electrons on the line joining the nuclei is usually zero. If there is an odd number of electrons, the projection of the total spin cannot be zero. Thus, a molecule of NO has spin 1/2. The spin of an O_2 molecule in the ground state is unity, which is an exception from the rule. A possible explanation of this is offered in [Sec. 34], where it is also shown that the projection of the orbital angular momentum on the axis joining the nuclei is zero

because the atomic electron shell of O_2 can be regarded as a deformed closed shell with two additional electrons. Their spins are parallel, and the zero projection of the orbital moment corresponds to the lowest energy level.

Disregarding the relatively few exceptions, we can write down the expression for the total energy of a diatomic molecule at ground state as the sum of three terms (see [34.16]):

$$\begin{aligned} \varepsilon &= \varepsilon_e + \varepsilon_{\text{vib}} + \varepsilon_{\text{rot}} \\ &= \varepsilon_e + h\omega_{\text{vib}} \left(v + \frac{1}{2} \right) + \frac{h^2 K (K+1)}{2mr_e^2} \end{aligned} \tag{3.2}$$

where K is the rotational quantum number of the molecule. Here, the last term is the smallest since it contains the mass of the nuclei in the denominator. Thus, $\varepsilon_e \sim 1/m^0$ (that is, does not depend on the nuclear mass m), $\varepsilon_{\text{vib}} \sim 1/m^{1/2}$, and $\varepsilon_{\text{rot}} \sim 1/m$.

Excitation of Electronic Levels. If we substitute the expression (3.2) into the Boltzmann distribution, the latter separates into the product of three distributions according to the electronic, rotational, and vibrational states. Let us suppose that a gas is at a temperature not exceeding 2000-3000 K. Then, if the energy of electronic excitation is several electron volts (recalling that 1 eV = 11 600 K), the fraction of molecules in excited electronic states, $e^{-\varepsilon_e/\theta}$, is very small. But as a rule dissociation of the molecules begins before any perceptible excitation of their electronic levels occurs.

Excitation of Vibrational Levels. Let us examine vibrational states. For generality we shall consider not only diatomic but polyatomic molecules as well. If their oscillations are harmonic we can, as was shown in [Sec. 7], go over to normal coordinates. Then the vibrational energy assumes the form of a sum of the energies of independent harmonic oscillators. The energy level for each oscillator is then given by a formula of the form (3.1) with a frequency ω_{vib} corresponding to a given normal oscillation.

Molecular oscillations may alter both the distances between neighbouring atoms and the angles between the "valence directions". For example, in a CO_2 molecule, which possesses a rectilinear equilibrium form O=C=O, there exist oscillations which change the distances between the O and C nuclei, as well as other oscillations that move the C nucleus out of the rectilinear configuration. The former type is called *valence oscillations*, and the second, *deformation oscillations*. The frequency of deformation oscillations is several times lower than that of valence oscillations. The estimate $h\omega_{\text{vib}} \sim \sim 0.1$ eV referred to valence oscillations. In compound normal oscillations of polyatomic molecules both types of nuclear displacement may occur.

In any case, if the vibrational energy separates into a sum of the energies of individual independent oscillations, then the distribution function also separates into a product of the distribution functions for each separate oscillation.

Let us express the mean energy per normal oscillation as

$$\bar{\varepsilon}_{\text{vib}} = \left\{ \sum_{v=0}^{\infty} h\omega_{\text{vib}} \left(v + \frac{1}{2}\right) \exp\left[-\frac{1}{\theta} h\omega_{\text{vib}} \left(v + \frac{1}{2}\right)\right]\right\}$$

$$\times \left\{ \sum_{v=0}^{\infty} \exp\left[-\frac{1}{\theta} h\omega_{\text{vib}} \left(v + \frac{1}{2}\right)\right]\right\}^{-1}$$

$$= \theta^2 \frac{\partial}{\partial \theta} \ln \left\{ \sum_{v=0}^{\infty} \exp\left[-\frac{1}{\theta} h\omega_{\text{vib}} \left(v + \frac{1}{2}\right)\right]\right\} \tag{3.3}$$

Transformation of the fraction to the derivative of the logarithm makes it possible to compute one sum instead of two. This device is constantly used in statistical physics. The sum in the lower line of (3.3) is called the *partition function.* It will be shown later on that the statistical properties of any system are determined by computing similar sums.

Equation (3.3) involves the partition function for a harmonic oscillator. It is easily computed. Indeed,

$$\sum_{v=0}^{\infty} \exp\left[-\frac{1}{\theta} h\omega_{\text{vib}} \left(v + \frac{1}{2}\right)\right] = e^{-h\omega_{\text{vib}}/2\theta} \sum_{v=0}^{\infty} \left(e^{-h\omega_{\text{vib}}/\theta}\right)^{v}$$

$$= \frac{e^{-h\omega_{\text{vib}}/2\theta}}{1 - e^{-h\omega_{\text{vib}}/\theta}} \tag{3.4}$$

Substituting this expression into (3.3) and differentiating, we obtain

$$\bar{\varepsilon}_{\text{vib}} = \frac{h\omega_{\text{vib}}}{2} + \frac{h\omega_{\text{vib}}}{e^{h\omega/\theta} - 1} \tag{3.5}$$

The first term in (3.5) denotes simply the zero energy of oscillation at a given frequency. The oscillation possesses this energy at absolute zero, because then the second term in (3.5) does not contribute anything. The second term has a very simple meaning. If we write the mean energy in terms of the mean vibrational quantum number $\bar{v}$

$$\bar{\varepsilon}_{\text{vib}} = \frac{1}{2} h\omega_{\text{vib}} + h\omega_{\text{vib}} \bar{v} \tag{3.6}$$

then

$$\bar{v} = \frac{1}{e^{h\omega_{\text{vib}}/\theta} - 1} \tag{3.7}$$

Hence the factor $(e^{h\omega_{\text{vib}}/\theta} - 1)^{-1}$ denotes the mean number of quanta an oscillation possesses at a temperature $\theta = k_B T$. At low temperatures $\bar{v}$ is close to zero. For example, for oxygen and nitrogen $h\omega_{\text{vib}}$ is about 0.2 eV, or 2000-3000 K. Therefore at room temperature oxygen and nitrogen are in their ground vibrational state. The reduced mass of a molecule of hydrogen is 14 times less than that of nitrogen. The energy of its vibrational quantum is close to 6000 K. In polyatomic molecules with deformation oscillations, such oscillations can be excited at temperatures of 300-600 K.

Vibrational Energy at High Temperatures. If the temperature is very high compared to $h\omega_{\text{vib}}$, the quantity $e^{h\omega_{\text{vib}}/\theta}$ can be replaced by $1 + h\omega_{\text{vib}}/\theta$. Substituting this into (3.5), we obtain

$$\bar{\varepsilon}_{\text{vib}} = h\omega_{\text{vib}}/2 + \theta \tag{3.8}$$

The first term does not relate to thermal excitation. Furthermore, it is considerably less than θ. We thus find that at sufficiently high temperature the mean energy per oscillation is equal to θ irrespective of the frequency. The same can be obtained proceeding from the nonquantized expression

$$\varepsilon_\omega = \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2} \tag{3.9}$$

for the energy of a harmonic oscillator. Substituting this into the Boltzmann distribution and calculating the mean energy, we obtain

$$\begin{aligned} \bar{\varepsilon}_\omega &= \left(\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, \varepsilon_\omega e^{-\varepsilon_\omega/\theta} \right) \left(\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, e^{-\varepsilon_\omega/\theta} \right)^{-1} \\ &= \theta^2 \frac{\partial}{\partial\theta} \ln \left(\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} dq\, e^{-\varepsilon_\omega/\theta} \right) \end{aligned} \tag{3.10}$$

The expression under the logarithm sign is called the *classical partition function.* It replaces the quantum partition function for quantities that change continuously and can easily be found with the help of known formulas (Exercise 2, Section 1):

$$\begin{aligned} \int_{-\infty}^{\infty} \exp\left(-\frac{p^2}{2m\theta} \right) dp \int_{-\infty}^{\infty} \exp\left(-\frac{m\omega^2 q^2}{2\theta} \right) dq &= (2\pi m\theta)^{1/2} \left(\frac{2\pi\theta}{m\omega^2} \right)^{1/2} \\ &= \frac{2\pi}{\omega}\theta \end{aligned} \tag{3.11}$$

Whence $\bar{\varepsilon}_\omega = \theta$. Neglecting the energy of zero oscillations, we find that the total vibrational energy of a gas for a frequency ω can be

written as follows:

$$\bar{E}_\omega = N_A\theta = RT \tag{3.12}$$

The contribution of this energy to specific heat is R. When $\theta \gg h\omega$, the specific heat due to vibrational degrees of freedom tends to a constant limit.

We shall now examine rotational energy.

Excitation of Rotational Levels.[4] The weight of a state with a given moment K is equal to $2K + 1$, corresponding to the number of possible projections of K. Of special interest is the case of a diatomic molecule consisting of two identical nuclei. In classifying the states of such a molecule it is necessary to take nuclear spin into account. Indeed, the wave equation for a molecule consisting of two identical nuclei does not change its form when the nuclei are interchanged. Hence, if the nuclei have half-integral spin, the wave function must be antisymmetric with respect to the interchange of the nuclei, and symmetric if the nuclear spin is integral or zero. The symmetry of the eigenfunction of a molecule is determined by the symmetry of its factors (in the approximation (3.2) it is separated into factors): electronic, vibrational, rotational, and nuclear spin. For most molecules, if the molecule is in its ground electronic state, the electronic term does not change with the interchange of the nuclei. The vibrational function depends only upon the absolute value of the distance between the nuclei, and therefore does not change either. The rotational eigenfunction is even with respect to this permutation in the case of even K, and odd in the case of odd K [Sec. 29]. Hence, if the nuclear spin is half-integral and the nuclei are subject to Pauli's exclusion principle, the spin function must be asymmetric in the case of even K and symmetric in the case of odd K. If the nuclear spin is integral and not zero, the reverse is true: the spin function is antisymmetric for odd K and symmetric for even K. And if the nuclear spin is zero, odd K are excluded, because then the spin factor of the wave function does not exist.

Rotational Energy of Para- and Ortho-Hydrogen. We shall now consider the rotational states of a hydrogen molecule. For hydrogen the total nuclear spin can be unity (the orthostate) or zero (the para-state) [33.42a, 33.42b]. The weight of a state with spin 1 is equal to 3, and with spin 0 it is 1. The state with $K = 0$ is even in the rotational wave function. Consequently, it must be odd

---

[4] The hypothesis that the rotational motion of molecules participates in the thermal motion of gases was advanced by M. V. Lomonosov in 1745.

in the spin function, that is, it must have spin 0. But the state with zero moment possesses the least rotational energy. Therefore, close to absolute zero hydrogen must be in the parastate.

At temperatures other than zero all states for which the Boltzmann factor $\exp\left[-h^2K(K+1)/(2mr_e^2\theta)\right]$ is of the order of unity are excited. Taking the moment of inertia of a hydrogen molecule equal to 0.45×10^{-40} g-cm², we see that at $T = 300$ K the summation over all odd moments,

$$\sum_{K=1,\,3,\,5,\,\ldots} (2K+1)\exp\left[-\frac{h^2K(K+1)}{2mr_e^2\theta}\right]$$

differs from the summation over even moments by several thousandths. But since for hydrogen the states with odd moments are ortho-states with respect to nuclear spin, each odd-moment state possesses an additional weight factor 3 according to the number of projections of spin 1. Consequently, at room temperature hydrogen comprises 3/4 ortho-hydrogen and 1/4 para-hydrogen. If hydrogen is cooled rapidly, the ratio 3 : 1 maintains for a long time, since the ortho-para transformation proceeds slowly. However, this is not the most probable state, since at absolute zero all molecules must have $K = 0$, which corresponds to the pure para-state.

One of the methods of obtaining pure para-hydrogen is to adsorb hydrogen on a substance that disrupts molecular bonds during adsorption, for example, activated charcoal. The hydrogen is then removed by reducing the pressure at low temperature; it changes to the para-state as it desorbs in the most probable state at the desorption temperature. If it is then heated to room temperature, it remains in the para-state for a fairly long time.

Let us now write the formulas for the mean rotational energy of ortho- and para-hydrogen. For simplicity we denote $h^2/(2mr_e^2)$ in the expression for rotational energy by the letter B. Then

$$\begin{aligned}\bar{\varepsilon}_{\text{para}} &= \left[\sum_{K=0,\,2,\,4,\,\ldots} (2K+1)\,e^{-BK(K+1)/\theta}\,BK(K+1)\right] \\ &\quad \times \left[\sum_{K=0,\,2,\,4,\,\ldots} (2K+1)\,e^{-BK(K+1)/\theta}\right]^{-1} \\ &= \theta^2\frac{\partial}{\partial\theta}\ln\left[\sum_{K=0,\,2,\,4,\,\ldots}(2K+1)\,e^{-BK(K+1)/\theta}\right] \qquad (3.13)\end{aligned}$$

For $\bar{\varepsilon}_{\text{ortho}}$ the summation is performed over odd K's. For a mixture at room temperature we have

$$\bar{\varepsilon}_{\text{rot}} = \frac{1}{4}\bar{\varepsilon}_{\text{para}} + \frac{3}{4}\bar{\varepsilon}_{\text{ortho}} \qquad (3.14)$$

At very low temperatures it is sufficient to retain only the term with $K = 2$, so that

$$\bar{\varepsilon}_{\text{para}} = \theta^2 \frac{\partial}{\partial\theta} \ln(1 + 5e^{-6B/\theta}) \approx 30\, Be^{-6B/\theta} \tag{3.15}$$

For ortho-hydrogen we obtain

$$\bar{\varepsilon}_{\text{ortho}} \approx \theta^2 \frac{\partial}{\partial\theta} \ln(3e^{-2B/\theta} + 7e^{-12B/\theta})$$
$$\approx 2B(1 + 14e^{-10B/\theta}) \tag{3.16}$$

Determination of Nuclear Spin from Rotational Specific Heat. The rotational specific heat of hydrogen can be used to determine proton spin. Consider Eq. (3.16). The first term is a constant. This means that even at absolute zero a molecule of ortho-hydrogen would have a rotational energy $2B$, which does not contribute to the specific heat since it does not depend on temperature. Defining specific heat as $\partial\bar{\varepsilon}/\partial\theta$, we find that for sufficiently low temperatures the ratio of the specific heat of ortho-hydrogen to that of para-hydrogen tends to zero, as the factor $e^{-4B/\theta}$. Consequently, if normal hydrogen is rapidly cooled to a low temperature, its rotational specific heat will be determined by the quarter of its molecules in the para-state. It will be one-fourth the rotational specific heat of pure para-hydrogen at the same temperature.

Thus, by measuring the specific heat of pure para-hydrogen and fast-cooled normal hydrogen, we can determine the spin of a proton or, knowing the spin from other data, we can prove that protons are subject to Pauli's exclusion principle since their wave function is antisymmetric.

Rotational Specific Heat of Molecules Consisting of Nonidentical Atoms. In diatomic molecules comprising nonidentical atoms the weight of states with respect to the nuclear spin is the same for odd and even K's. Therefore, their mean rotational energy is expressed in the form

$$\bar{\varepsilon}_{\text{rot}} = \theta^2 \frac{\partial}{\partial\theta} \left[\ln \sum_{K=0}^{\infty} (2K+1)\, e^{-BK(K+1)/\theta} \right] \tag{3.17}$$

The sum cannot be written in finite form but is easily tabulated numerically. Let us evaluate the temperature at which we can justifiably go over from the summation to an integral. For hydrogen

$$B = \frac{h^2}{2mr_e^2} \approx \frac{1.11 \times 10^{-54}}{1.67 \times 10^{-24} \times (0{\cdot}74)^2 \times 10^{-16}} = 1.2 \times 10^{-14}\,\text{erg}$$

The ratio B/θ is of the order of unity at $\theta = 87$ K.

In this estimate m is the reduced mass of two protons equal to half the proton mass, and $r_e = 0.74 \times 10^{-8}$ cm, which corresponds to a moment of inertia of approximately 0.45×10^{-40} g-cm^2. For other gases B is of the order of several kelvins, so that at temperatures for which these gases are not in the liquid state the ratio B/θ is a small quantity. Accordingly, the summation in (3.17) may be replaced by an integral. If we denote $K(K+1) = x$, then $(2K+1)\,dK = dx$ and

$$\sum_{K=1}^{\infty} (2K+1)\, e^{-B(K+1)K/\theta} \approx \int_0^{\infty} e^{-Bx/\theta}\, dx = \frac{\theta}{B} \tag{3.18}$$

Substituting this into (3.17), we obtain for the rotational energy of a diatomic molecule or any linear molecule the formula

$$\bar{\varepsilon}_{\mathrm{rot}} = \theta = \frac{RT}{N_{\mathrm{A}}} \tag{3.19}$$

Note that the concepts of high temperature for oscillations and rotations do not coincide. For the rotational specific heat of oxygen, the temperature must be above 10 K to be regarded as high, but for the vibrational specific heat high temperature is above 2000 K. That is why the specific heat of diatomic gases is constant within a wide temperature range (notably, at room temperature) and consists of a translational portion $3R/2$ and rotational portion R, so that the total specific heat is $5R/2$.

Note that rotational specific heat does not tend to R monotonically and passes through a maximum of $1.1R$ at $\theta = 0.81B$.

The rotational energy of polyatomic molecules will be discussed in Section 8.

EXERCISES

1. The deformational vibration of a linear symmetrical triatomic molecule ABA consists in a displacement of atom B in one direction perpendicular to the line ABA and the displacements of both A atoms to equal distances in the opposite direction. Such a configuration of displacements precludes any rotation of the molecule in space. The displacements must be such that the molecule's centre of mass remains stationary. The vibrations may be in two mutually perpendicular planes. The masses of the atoms, the distance AB, and the vibration frequency are given. Determine the mean square of the angle between lines AB and BA, that is, the angle of deviation of the molecule from its rectilinear form.

Solution. Since the A atoms undergo equal displacements in the deformational vibration, the reduced mass for this type of vibration is

$$m = \frac{2m_A m_B}{2m_A + m_B}$$

The potential energy of the vibration is equal to

$$\frac{m}{2} l^2 \omega^2 \varphi^2$$

where l is the distance AB, and φ is the angle between AB and BA ($\varphi \gg 1$).

The mean value of the potential energy is equal to half the mean value of the total energy. The total energy of deformational vibration consists of the vibration energies of equal frequencies in mutually perpendicular planes, that is, it is equal to

$$h\omega\left(v_1 + \frac{1}{2}\right) + h\omega\left(v_2' + \frac{1}{2}\right) = h\omega(v_1 + v_2 + 1)$$

where v_1 and v_2 are any integers or zero. Consequently,

$$\frac{m}{2} l^2 \omega^2 \overline{\varphi^2} = \frac{h\omega}{2}(\bar{v}_1 + \bar{v}_2 + 1), \qquad \overline{\varphi^2} = \frac{h}{ml^2\omega}\left(\frac{2}{e^{h\omega/\theta} - 1} + 1\right)$$

2. Determine the rotational energy of para- and ortho-deuterium.

Solution. Particles with spin 1 have a symmetric wave function. The projection of spin 1 assumes three values: 1, 0, —1. Denoting the spin wave functions of both deuterons corresponding to these projections as $\psi_1(1)$, $\psi_1(0)$, $\psi_1(-1)$ and $\psi_2(1)$, $\psi_2(0)$, $\psi_2(-1)$, let us form all the spin wave functions of deuterium corresponding to the total projection of spin 0.

Taking only symmetric

$$\psi_1(1)\psi_2(-1) + \psi_1(-1)\psi_2(1), \qquad \psi_1(0)\psi_2(0)$$

and antisymmetric

$$\psi_1(1)\psi_2(-1) - \psi_2(1)\psi_1(-1)$$

combinations, we get for the total projection of spin ± 1,

$$\psi_1(1)\psi_2(0) + \psi_1(0)\psi_2(1), \qquad \psi_1(1)\psi_2(0) - \psi_1(0)\psi_2(1)$$
$$\psi_1(-1)\psi_2(0) + \psi_1(0)\psi_2(-1), \qquad \psi_1(-1)\psi_2(0) - \psi_1(0)\psi_2(-1)$$

and for the total projection of spin ± 2,

$$\psi_1(1)\psi_2(1), \qquad \psi_1(-1)\psi_2(-1)$$

The symmetric state has the maximum spin projection of 2. Hence, all wave functions with total spin 2 constituting one state are symmetric. There are five such functions (according to the number of projections of spin 2). There are three antisymmetric functions, which corresponds to the number of projections of spin 1. Consequently, the function with zero spin projection is symmetric, since the total number of functions is ten. Note that here we constructed the eigenfunctions of the operator of spin projection, not its absolute value. But for the problem concerning us it is sufficient to know the number of states and not the exact form of their wave functions.

Thus, deuterium has six ortho-states with a symmetric spin function and three para-states with an antisymmetric spin function. A rotational wave function with even K's corresponds to the former, and a rotational wave function with odd K's to the latter. Then the total wave function is symmetric, as should be in the case of integral deuteron spin. The weight of states due to spin is six for the ortho-states and three for the para-states. Therefore the statistical sum of ortho-deuterium is

$$6 \sum_{K=0,\,2,\,4,\,\ldots} (2K+1) \exp\left(-\frac{BK(K+1)}{\theta}\right)$$

and for para-deuterium it is

$$3 \sum_{K=1,\,3,\,5,\,\ldots} (2K+1) \exp\left(-\frac{BK(K+1)}{\theta}\right)$$

At absolute zero deuterium must occur in its ortho-state. The energies of both states are, from Eqs. (3.15) and (3.16) respectively,

$$\varepsilon_{\text{ortho}} \approx 30e^{-6B/\theta}$$

$$\varepsilon_{\text{para}} \approx B\left(2+28e^{-10B/\theta}\right)$$

Compared with hydrogen, the ortho- and para-states are interchanged. Close to absolute zero the ortho-state makes the major contribution to specific heat. At room temperature, 2/3 the molecules are in the ortho-state. That is why the rotational specific heat of fast-cooled deuterium is less than that of deuterium prepared by low-temperature desorption (their proportion is 2/3). By measuring this ratio it can be shown that the spin of a deuteron is unity, not zero. Note that in the second case the only states were with even K's, that is, only ortho-states (at any temperature).

4

APPLICATIONS OF STATISTICS TO ELECTROMAGNETIC FIELDS IN VACUUM AND TO CRYSTALLINE BODIES

The Most Probable State in a System Comprising Matter and Radiation. Imagine a closed cavity in an opaque body. The walls of the cavity are capable of absorbing and emitting electromagnetic radiation. The interrelationship of emission and absorption follows from the fact that when a direct process is permissible, the reverse process is also permissible (see [Sec. 36]). By opacity of the walls is meant

that they absorb radiation of all frequencies and hence can emit radiation of all frequencies. It is therefore possible for the most probable state to develop in the cavity, at which the same radiation energy is absorbed and emitted by unit surface in unit time in all directions.

The most probable radiation state is steady in the same sense as the most probable state of a gas examined in the preceding sections. The important thing is that the radiations is characterized by a specific temperature equal to the temperature of the walls. The necessity of equal temperatures will be shown later on in setting forth the fundamentals of thermodynamics (Secs. 7 and 8). For the present let us accept it as an assumption.

Black Body. Radiation in a cavity can be studied experimentally by making a small aperture in its wall. If it is sufficiently small, it does not affect the state within the cavity. The radiation impinging on such an aperture from the outside is absorbed on the inside and does not leave the cavity. In this sense the aperture is like a black body which does not reflect light rays. That is why it is called a *black body*, and the radiation coming out from the aperture is called *black body radiation*. Obviously, the escaping radiation is that which was inside. The radiation incident from outside is scattered in repeated reflections from the walls, and the greater part is absorbed in each reflection. The amount of energy leaving the cavity is extremely small.

The term "black body" is rather paradoxical, as it contradics the obvious picture. Actually, though, a black body radiates more than a nonblack body at the same temperature because it absorbs more, and in the most probable state emission and absorption are equal. If a body with a cavity and an aperture is heated to a glow, the aperture will exhibit the brightest glow.

Statistics of an Oscillator Representation of a Field. The Planck Radiation Formula. Let us consider the applications of statistics to black body radiation. For this it is necessary to quantize the electromagnetic field, as was done in [Sec. 36]. Unlike the statistics of a gas, the statistics of radiation does not permit a transition to classical equations from which the action quantum is eliminated entirely. This will become clear a little later.

In quantizing a field, the wave or corpuscular approach is possible. The field is represented, as was done in [Sec. 36], by a set of linear harmonic oscillators, each of which is characterized by its wave vector **k** and polarization σ. Obviously, the oscillators differ in the values of **k** and σ. Their quantum properties are apparent not in calculating the number of states of the field but in the fact that the energy of each one cannot be an arbitrary number and belongs

to a discrete energy spectrum [27.23, 27.28*b*], that is, is equal to $h\omega(n + 1/2)$, where n is an integer. This is the characteristic case of a Boltzmann gas (see Secs. 2, 3).

In the most probable state of an oscillator, the mean number $\bar{n}$ of vibrational quanta $h\omega$ is given by a formula of the same form as (3.7):

$$\bar{n} = \frac{1}{e^{h\omega/\theta} - 1} \tag{4.1}$$

The number of oscillations with frequency ω, according to Eq. (1.31), is given by the formula

$$dg(\omega) = \frac{V\omega^2\, d\omega}{\pi^2 c^3} \tag{4.2}$$

Here both possible field polarizations are taken into account and also $\omega = k/c$. Thus the energy of an electromagnetic field in a frequency interval $d\omega$ is

$$dE(\omega) = \bar{n}\, dg(\omega) = \frac{V\omega^3 h\, d\omega}{\pi^2 c^3 (e^{h/\omega\theta} - 1)} \tag{4.3}$$

The radiation spectrum of the sun is close to this frequency distribution.

Statistics of Light Quanta. Equation (4.1) may be obtained in another way. Applying quantum mechanics to separate oscillators, we can represent an electromagnetic field as an assembly of elementary particles called *light quanta*. Quanta of the same frequency, line of propagation, and polarization are indistinguishable from one another. They possess integral angular momentum eigenvalues. This was mentioned in [Sec. 36].[5] Therefore, they are not subject to Pauli's exclusion principle and satisfy the Bose-Einstein statistics. However, unlike the molecules of gases subject to the Bose distribution, the number of quanta is not constant, as they are absorbed and emitted. The additional condition (1.12) does not hold for them.

It is easy to go over from the general Bose distribution to the special case when condition (1.12) does not apply. For this it is sufficient to assume the parameter μ, by which Eq. (1.12) is multiplied and introduced to satisfy the conditions N = const, to be zero. Then the Bose distribution is simplified thus:

$$\bar{n} = \frac{1}{e^{\varepsilon/\theta} - 1} \tag{4.4}$$

---

[5] Generally speaking, an eigenmoment of unity has three projections. But a zero projection would correspond to a longitudinal wave, which does not exist. Two circular polarizations correspond to two projections of the angular momentum on the wave vector, the projections being ± 1 (cf. [Sec. 18]).

Taking into account that for a quantum $\varepsilon = h\omega$, we again obtain (4.1). Thus, formula (4.1) denotes both the mean vibrational quantum number of an oscillator in an assembly subject to Boltzmann statistics and the mean number of vibrational quanta. We distinguish between quantum and nonquantum statistics, it will be recalled, precisely according to whether the particles are distinguishable or not (Sec. 2).

The Impossibility of the Limiting Transition $h \to 0$ in the Statistics of an Electromagnetic Field. Let us now turn to the oscillator picture. According to the classical theory, the mean energy of an oscillator is equal to θ (see Eq. (3.12)). If we multiply it by $g(\omega)$, we obtain the classical *Rayleigh-Jeans formula* for the energy of black body radiation:

$$dE(\omega)_{\text{class}} = \frac{V\omega^2\theta\, d\omega}{\pi^2 c^3} \tag{4.5}$$

But this formula is obviously inadequate at high frequencies: upon integration over all frequencies it yields an infinite total energy of the field. It was precisely here that the limitations of classical notions in statistics were revealed most convincingly.

Proceeding from experimental data concerning energy distribution in the spectrum of black body radiation, Max Planck in 1900 proposed Eq. (4.3). It was here that the quantum of action appeared for the first time in physics.

Equation (4.5) holds only for small frequencies, that is, when $h\omega \ll \theta$.

Black Body Radiation. The total energy of equilibrium electromagnetic radiation can be found without difficulty from formula (4.3). Integrating with respect to ω, we obtain

$$E = \frac{Vh}{\pi^2 c^3} \int_0^\infty \frac{\omega^3\, d\omega}{e^{h\omega/\theta} - 1} = \frac{V}{\pi^2 c^3} \frac{\theta^4}{h^3} \int_0^\infty \frac{x^3\, dx}{e^x - 1} \tag{4.6}$$

The integral in (4.6) is a dimensionless number equal to $\pi^4/15$, and the energy is proportional to the fourth power of the temperature.

The result (4.6) can be verified from the radiant emittance of a black body. It is easy to relate it to E. For this it is sufficient to calculate the number of quanta falling from inside in unit time upon unit surface of a cavity, normal to the surface. If a small aperture is made in the wall, radiation of the same composition as the incident radiation will pass through it.

The velocity of each quantum is c, hence its normal component is equal to $c \cos \vartheta$, where ϑ is the angle with the normal. If the quanta fall on the surface at an angle ϑ, in unit time all the quanta

within a cylinder of height $c \cos\vartheta$ constructed on unit surface (that is, having unit base) will strike that part of the surface. The energy within the cylinder is $(E/V)\, c \cos\vartheta$. The proportion of quanta travelling at angle ϑ to the surface is

$$\frac{1}{4\pi}\int_0^{2\pi} d\varphi \sin\vartheta\, d\vartheta = \frac{1}{2}\sin\vartheta\, d\vartheta$$

From this we obtain the expression for the total energy flux in unit time at any angle ϑ:

$$\frac{1}{2}\int_0^{\pi/2} \sin\vartheta\, d\vartheta \cos\vartheta \frac{cE}{V} = \frac{c}{4}\frac{E}{V} = \frac{\pi^2\theta^4}{60c^2h^3} = \frac{\pi^2 k_B^4 T^4}{60c^2h^3} \tag{4.7}$$

The constant in front of T^4 is equal to 5.67×10^{-5} erg-cm^{-2}K^{-4}s^{-1}. Equation (4.7) cannot be applied directly to heated solid bodies without ascertaining the extent to which they can be regarded as black.

Because the sun's chromosphere (its luminous layer) is nearly opaque to radiation, its emission spectrum closely approximates the spectrum of a black body. The radiation temperature is close to 5750 K.

Pressure of Black Body Radiation. The pressure of black body radiation is easily calculated. For this it is convenient to make use of the reasoning that led to Eq. (4.7). Only now we must compute not the number of quanta but the normal component of their momentum carried across unit area. This component is equal to the energy of a quantum divided by c and multiplied by $\cos\vartheta$. Therefore, unlike the procedure in deriving (4.7), we must now integrate not $\cos\vartheta$ but $\cos^2\vartheta$. Furthermore, to every quantum arriving at the surface at the most probable state of the field there is a similar quantum radiated by the wall in the opposite direction. Therefore the transferred momentum is doubled. Consequently,

$$p = \frac{E}{cV}\frac{2c}{2}\int_0^{\pi/2} \cos^2\vartheta \sin\vartheta\, d\vartheta = \frac{E}{3V} \tag{4.8}$$

This means that the pressure is equal to 1/3 the energy density. Equation (2.22) yields the same result if the momentum is taken equal to ε/c instead of mv. It should be noted that in P. N. Lebedev's experiments, in which the pressure of a directed beam was measured instead of radiation arriving uniformly from all directions, it was found that $p = E/V$, that is, the pressure is equal to the energy density (without the factor 1/3).

The pressure of electromagnetic radiation, according to (4.8) and (4.6), increases in proportion to the fourth power of the temperature, whereas gas pressure is, roughly speaking, proportional to the first power of temperature. Hence, at sufficiently high temperatures radiation pressure always predominates.

At high temperatures the pressure of a substance can be calculated according to the ideal-gas formula, since the interaction energy of the particles becomes small in comparison with their kinetic energy. Consequently,

$$p_s = \frac{N_A \theta}{V}$$

If we assume that the atoms have dissociated into nuclei and electrons, there is no difficulty calculating the ratio of N_A/V to the density of the substance. Suppose that the substance is hydrogen. Then to every proton there is one electron. If ρ is the density of the substance, then the ratio N_A/V is equal to $2\rho/m$, where m is the mass of the proton, and the factor 2 takes into account the electron. Hence,

$$p_s = \frac{2\rho}{m}\theta \tag{4.9}$$

According to (4.6) and (4.8), the radiation pressure is

$$p_r = \frac{\pi^2}{45\,(hc)^3}\theta^4 \tag{4.10}$$

From this we obtain the density-temperature relationship for the case when radiation pressure equals gas pressure:

$$\rho = \frac{\pi^2}{90}\frac{m\theta^3}{(hc)^3} = 1.5 \times 10^{-23}\frac{\mathrm{g}}{\mathrm{cm^3\,K^3}}T^3$$

For example, at $\rho = 1$ g-cm^{-3} both pressures are equal at 4×10^7 K. At higher temperatures the radiation pressure predominates. It substantially affects processes taking place in the interiors of certain classes of stars.

The Maximum of the Black Body Radiation Spectrum. The maximum energy per unit frequency interval occurs at the frequency given by the equation

$$\frac{d}{d\omega}\left[\omega^3 \left(e^{h\omega/\theta} - 1\right)^{-1}\right]_{\omega=\omega_{\max}} = 0 \tag{4.11}$$

which yields

$$1 - \exp\left(-h\omega_{\max}/\theta\right) = h\omega_{\max}/\theta \tag{4.12}$$

This equation has a single solution

$$h\omega_{\max}/\theta = 2.822 \tag{4.13}$$

Hence, the frequency corresponding to the maximum in the black body radiation spectrum is directly proportional to the absolute temperature (*Wien's displacement law*)

$$\omega_{\max} = \frac{2.822}{h}\theta \tag{4.14}$$

Note that the numerical coefficient would have been different if we considered the wavelength distribution instead of the frequency distribution. It is interesting to note that the corresponding wave-

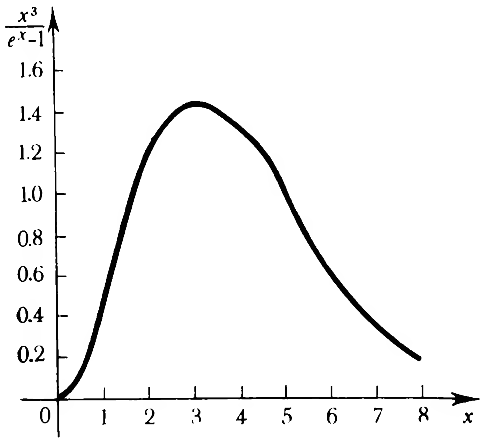


Figure 3

length $\lambda_{\max}$ in the solar spectrum is very close to the maximum sensitivity region of the human eye. The distribution of $(h\omega/\theta)^3 \times$ $\times$ $(e^{h\omega/\theta} - 1)^{-1}$ is shown in Figure 3. If instead of $x = h\omega/\theta$ we plot the frequency ω along the abscissa, the maxima lie on the straight line $\omega = 2.822\theta/h$, which passes through the origin.

Spontaneous and Stimulated Emission of Quanta. It was shown at the beginning of this section that in order to attain the most probable state radiation quanta must be emitted and absorbed. Radiation totally isolated from matter can be represented as an assembly of noninteracting linear oscillators incapable of exchanging energy. Any initial energy distribution among them would remain unchanged. Apparently, radiation inside a cavity with ideally reflecting walls could attain the most probable distribution if we mentally placed a "speck of carbon" inside the cavity. Small enough not to distort the radiation field, it would at the same time redistribute the energy among the oscillators by emitting and absorbing quanta.

The probability of quantum emission and absorption was calculated in [Sec. 36]. If in a field there are present n quanta with a given wave vector $\mathbf{k}$ and polarization σ, the probability of the emission of one more such quantum is proportional to $n + 1$, and the probability of absorption is proportional to n [Eq. (36.25) ff.]. The square of the modulus of the matrix element, being the coefficient of $n + 1$ or n, is the same. Emission proportional to n is called *stimulated* or *induced.* It was predicted by Albert Einstein in 1916. Emission independent of n is called *spontaneous.*

For large values of n stimulated emission is stronger than spontaneous emission. But large n corresponds to the transition to the classical approximation in electrodynamics [Sec. 36]. That is why stimulated emission is in essence a classical effect, which can be visualized as follows. An electromagnetic wave incident on a system vibrates the radiating particles. The vibration amplitude is proportional to the wave amplitude, hence the radiation intensity is proportional to the square of the amplitude of the incident wave, or to the number of quanta, n. Spontaneous emission is a purely quantum effect.

Let us now consider the most probable state that forms in a system comprising atoms and an electromagnetic field. As will be explained later on, it is natural to call this state *thermal equilibrium* (Sec. 8). We shall show that the Planck distribution (4.1) is established in the system.

Let us denote the energies of two atomic states by symbols ε_1 and ε_0. In the transition from state 1 to state 0 quanta with energy $\hbar\omega = \varepsilon_1 - \varepsilon_0$ are emitted. In the reverse transition they are absorbed.

The number of acts of absorption by all N_0 atoms in state 0 in unit time is equal to

$$N_0 W_{01} n \tag{4.15a}$$

where W_{01} is a factor of proportionality.

The number of acts of emission by all N_1 atoms in state 1 per unit time (taking into account both spontaneous and stimulated acts) is

$$N_1 W_{10} (n+1) \tag{4.15b}$$

The equilibrium condition consists in the equality of the quantities (4.15*a*) and (4.15*b*):

$$N_0 W_{01} n = N_1 W_{10} (n+1) \tag{4.16}$$

Here, we substitute N_0 and N_1 from Eq. (2.1):

$$e^{(\mu-\varepsilon_0)/\theta} g_0 W_{01} n = e^{(\mu-\varepsilon_1)/\theta} g_1 W_{10} (n+1) \tag{4.17}$$

According to Eqs. [32.42] and [36.26], the factor of proportionality W_{01} is the square of the modulus of the matrix element of the transition (calculated according to the atom's state) multiplied by the weight of the atom's state after the transition, g_1 (the meaning of W_{10} is similar). Therefore, the factors $g_0 W_{01}$ and $g_1 W_{10}$ in Eq. (4.17) cancel out. Besides, $\varepsilon_1 - \varepsilon_0 = h\omega$, which gives

$$e^{h\omega/\theta} n = n + 1 \tag{4.18}$$

whence the Planck distribution is immediately obtained:

$$n = \frac{1}{e^{h\omega/\theta} - 1} \tag{4.19}$$

Note that the Bose-particles field theory always leads to the concept of stimulated emission. The probability of the appearance of the $(n+1)$st particle in the field is proportional to $n+1$, and the probability of the disappearance is proportional to n. This follows from the quantized field equations describing bosons, similar to what was obtained in [Sec. 36] for an electromagnetic field. If the Bose particles possess a charge, as for example $\pi^{\pm}$ mesons, only transitions compatible with the conservation of the total charge of the system are possible.

With Fermi particles a transition to a filled level is impossible. Therefore fermion fields are quantized in such a way that the transition probability includes the factor $1 - f$ if the probability that the given level is filled is equal to f. This is achieved by appropriate antisymmetrization of the wave function of the fermion system (see [Sec. 33]).

Lasers. If $\varepsilon_1 > \varepsilon_0$ and the atoms occupy a Boltzmann energy distribution, the number of atoms in state 1 is smaller than in state 0. But by outside action on the atoms it is possible to have more of them at the higher level than at the lower one. In such cases we speak of *population inversion* of the system (meaning the population or occupation of energy levels). If, for example, the radiation transition $1 \to 0$ is prohibited (see [Sec. 36]), it is possible in one way or another to accumulate a large number of atoms in state 1, for instance, by electron collisions. But if the $1 \to 0$ transition does occur, every emitted quantum induces further transitions according to the stimulated emission mechanism. Since in the process the quanta are emitted in the same direction and with the same polarization, it produces a powerful peak of coherent, one-directional radiation. That is the principle on which a laser operates.

For more than one hundred years it was thought that it is impossible to induce different atoms to emit coherent radiation. Although stimulated emission is, as mentioned, a classical effect by nature,

it had to be formulated in quantum terms to arrive at the idea of a laser, which was done by V. A. Fabrikant in 1940.

The Oscillation Spectrum of a Solid-Body Lattice. The statistical behaviour of the crystalline lattice of a solid body in many ways resembles the behaviour of an electromagnetic field. Therefore, before applying statistics to lattice oscillations, they must be represented, as far as possible, in the same form as the field oscillations described in [Sec. 36].

The meaning of this representation consists in reducing the oscillations to normal coordinates. As applied to radiation, it resulted in expressing the electromagnetic field in the form of a superposition of travelling plane waves. The amplitude of each wave was in fact a normal coordinate of the field. Normal oscillations of atoms in a lattice are also the amplitudes of waves travelling through the lattice from one atom to another (that is, through the discrete assembly of atoms). In the linear (harmonic) approximation, such oscillations in the form of travelling waves are mutually independent, and the total energy is the sum of the energies of individual oscillators. An example of such a wave will be given in Exercise 4 at the end of this section.

However, the following differences exist between an assembly of oscillators for an electromagnetic field and for a solid crystalline body.

(1) The number of degrees of freedom of an electromagnetic field is infinite, so that it contains all frequencies from 0 to ∞. A solid body has a finite number of degrees of freedom equal to $3N$, where N is the number of atoms. Therefore the vibration frequencies range from zero to some maximum frequency ω_{max}.

(2) The frequency dependence of the wave vector of an electromagnetic field is given by the simple law $\omega = ck$. In solid-body oscillations, the frequency depends on the wave vector in a complex manner, and this dependence differs for different crystals. Only in the limit, for very long wavelengths (small k's), do the atomic vibrations become vibrations of a continuous medium according to the laws of elasticity theory. For such oscillations frequency is proportional to the wave vector, and the atomic structure of the crystal can be neglected.

(3) The situation is more complicated when a crystal consists of atoms of different types or of atoms occupying positions that are displaced in relation to the equilibrium positions in a crystal. In that case there exist oscillations in which neighbouring atoms move in different phases even when the length of the wave travelling through the crystal is large compared to the period of the crystal. Such oscillations do not turn into oscillations of an elastic continuum. Figure 4 presents oscillations of two types shown, for the

sake of simplicity, not in a lattice but in a one-dimensional chain made up of two types of atoms. In case (*a*), the "white" and "black" atoms displace rectilinearly in the same direction from the equilibrium position. In the limit, when the wavelength tends to infinity, the atomic structure of the chain does not affect waves of this type. In case (*b*), the displacements are in opposite directions. At infinite wavelength the restoring force is greatest, and the frequency corresponding to the vibrations is maximum, not zero.

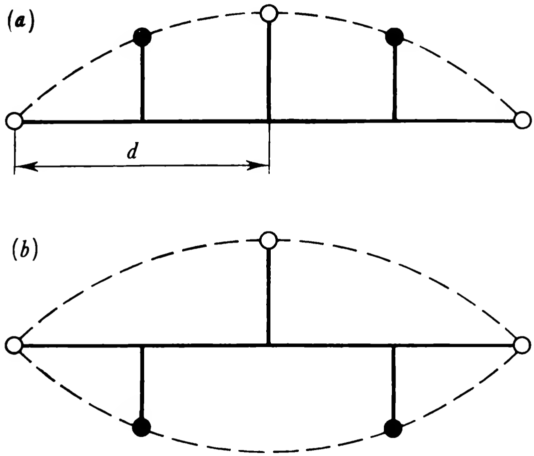


Figure 4

If the lattice (or chain) is ionic, the atoms carry opposite charges. A wave of type (*b*) leads apparently to the appearance of a dipole moment (see [Sec. 16]); hence the oscillation interacts strongly with the electromagnetic field. It is therefore termed *optical oscillation.* Type (*a*) is called *acoustic*, since in the limit of long wavelengths it corresponds to an elastic, that is acoustic, wave in a continuous medium.

If there are i atoms in an elementary cell of a crystal, then in the three-dimensional case $3i$ types of oscillations occur. Three of them are acoustic, corresponding to case (*a*), the rest correspond to type (*b*). For instance, every cell of a diamond lattice contains two carbon atoms. In this case there are three types of oscillations and three b-type oscillations.

The total number of oscillations of a crystal is $3iN' = 3N$, where N' is the number of elementary cells. It is obvious that the number of normal oscillations equals the number of degrees of freedom, that is, three times the number of atoms in the lattice.

Calculation of the number of oscillations per certain interval of values of the wave vector is carried out in the same way as for an electromagnetic field. Namely, the periodicity condition must be superimposed, assuming that the wave picture is reproduced exactly in the displacements a_1, a_2, and a_3 along all three coordinate axes (see [Sec. 36]). Then the wave vector passesses components proportional to integers n_1, n_2, n_3:

$$k_x = \frac{2\pi n_1}{a_1}, \qquad k_2 = \frac{2\pi n_2}{a_2}, \qquad k_3 = \frac{2\pi n_3}{a_3} \tag{4.20}$$

To each triad of numbers corresponds one oscillation of a specific type out of the total number of $3i$. To some interval $d\mathbf{k}$ there corresponds a number $dg(\mathbf{k})$ equal to

$$dg(\mathbf{k}) = dn_1\, dn_2\, dn_3 = \frac{a_1 a_2 a_3\, dk_x\, dk_y\, dk_z}{(2\pi)^3} = \frac{V\, dk_x\, dk_y\, dk_z}{(2\pi)^3} \tag{4.21}$$

The Energy of a Solid Body. It is now easy to write the expression for the amount of energy within the interval $d\mathbf{k}$. If we denote the type of oscillation by the letter σ (by analogy with the polarization of electromagnetic waves), to the wave vector $\mathbf{k}$ correspond a frequency ω_σ and energy quantum $h\omega_\sigma$. From this we obtain

$$dE_\sigma(\mathbf{k}) = \frac{V\, dk_x\, dk_y\, dk_z\, h\omega_\sigma}{(2\pi)^3 (e^{h\omega_\sigma/\theta} - 1)} \tag{4.22}$$

To find the total energy of the crystal we must integrate the expression over $dk_x\, dk_y\, dk_z$ and sum over σ. Unlike the procedure in the case of an electromagnetic field (cf. Eq. (4.6)), here the integration must be performed not to infinity but only between limits such that the total number of oscillations equals the total number of degrees of freedom $3N$:

$$V \sum_\sigma \iiint \frac{dk_x\, dk_y\, dk_z}{(2\pi)^3} = 3N \tag{4.23}$$

To each value of σ corresponds a specific function $\omega_\sigma(\mathbf{k})$, called also an *oscillation branch*. For small k's there is a simple dependence for the three acoustic branches:

$$\omega_\sigma = c_\sigma k$$

where the velocity of propagation c_σ depends only upon the direction but not the value of $\mathbf{k}$. But for large k's the curve is not monotonic and has a complex form. Optical branches have both maximum and minimum frequencies, usually of compatible value.

The expression (4.22) cannot be integrated in general form if only because there is no general dependence $\omega_\sigma(\mathbf{k})$ for all crystals. There

are, however, two important limiting cases when it is possible to obtain a general form of the energy-temperature dependence.

(1) The temperature θ is considerably greater than the limiting-frequency quantum $h\omega_{max}$. It is then all the more greater for all other quanta, and we can restrict ourselves to the first term of the series expansion of the exponential function:

$$\exp\left(\frac{h\omega_\sigma}{\theta}\right) \approx 1 + \frac{h\omega_\sigma}{\theta}$$

Substituting this into (4.22), we obtain a very simple expression for the lattice energy:

$$E = V\theta \sum_\sigma \int\int\int \frac{dk_x\, dk_y\, dk_z}{(2\pi)^3} = 3N_A\theta = 3RT \tag{4.24}$$

Here we have made use of the condition (4.23). The meaning of Eq. (4.24) is very simple: the number of oscillators equals the number of degrees of freedom, and at high temperatures the energy of each oscillator is, from Eq. (3.12), equal to θ. The specific heat of the lattice is $3R$ and the same for all elements in molar units. This law holds well for many elements already at room temperatures (the *Dulong and Petit law*). Exceptions are, for example, diamond and beryllium, for which the large frequency ω_{max} is due to a relatively small atomic weight, since the oscillation frequency is proportional to $M^{-1/2}$.

The Dulong and Petit law holds poorly for crystals consisting of separate molecules. Such crystals have very many oscillation branches. Some approximate oscillations of the molecules as a whole, others approximate intramolecular oscillations. For such branches the inequality $h\omega_{max} \ll \theta$ does not hold at room temperatures, while at higher temperatures molecular crystals usually liquefy or sublimate.

(2) The temperature θ is considerably less than $h\omega_{max}$. Then the factor $(e^{h\omega/\theta} - 1)^{-1}$ is so small that integration can be taken to infinity without any substantial error. Only small frequencies, with quanta of the order of θ, than is, $h\omega \approx \theta$, make any noticeable contribution. For large frequencies the Planck factor $(e^{h\omega/\theta} - 1)^{-1}$ cancels out the contributions of the corresponding oscillations. Optical branches do not participate in thermal excitations at all since their minimum frequencies correspond to quanta that are considerably greater than θ.

There remain only longwave oscillations of acoustic branches. They correspond to the vibrations of a continuous medium whose frequency is linked with the wave vector by the ralationship mentioned before:

$$\omega_\sigma = c_\sigma\left(\frac{\mathbf{k}}{k}\right) k \tag{4.25}$$

It is expedient to transform the volume element $dk_x \, dk_y \, dk_z$ to spherical coordinates, that is, replace it by the expression $k^2 \, dk \, d\Omega$, where $d\Omega$ is the solid-angle element for the directions of $\mathbf{k}$. Then the integration with respect to k must, in accordance with what has just been said, be taken from 0 to infinity.

Thus we obtain the following formula for the total energy of a crystal at low temperatures:

$$E = \frac{Vh}{(2\pi)^3} \sum_{\sigma=1}^{3} c_\sigma \, d\Omega \int_0^\infty \frac{k^3 \, dk}{e^{hc_\sigma k/\theta} - 1} \tag{4.26}$$

The integral is found in the same way as in the calculation of the energy of an electromagnetic field according to Eq. (4.6).

Consequently

$$E = \frac{\pi V}{120} \frac{\theta^4}{h^3} \int \left(\sum_\sigma \frac{1}{c_\sigma^3} \right) d\Omega \tag{4.27}$$

The energy of a crystal lattice at low temperatures, like the energy of an electromagnetic field, is proportional to the fourth power of the temperature. Specific heat is proportional to the third power.

The Debye Interpolation Formula. Peter J. W. Debye, who enunciated the theory of specific heat of crystals at low temperatures, proposed an interpolation formula for intermediate temperatures, when formulas (4.24) and (4.27) do not hold. The Debye equation reduces to both these formulas in the limiting cases of high and low temperatures. The intermediate interval is described qualitatively, but in certain agreement with experience. For this we assume that the law

$$\omega_\sigma = c_\sigma k$$

holds for all k's (not only small values). But after such a major simplification there is no sense in taking into account that the velocity of elastic waves, c_σ, depends on the direction of vector $\mathbf{k}$. Waves in that case should be considered not in a crystalline continuum but simply in an isotropic elastic body. Waves can travel through it transversely with a velocity c_t, and longitudinally with a velocity c_l. Obviously, the transverse waves have two polarizations ($\sigma = 1, 2$). We denote the polarization of longitudinal waves by the index $\sigma = 3$. Owing to isotropy of the body, c_t and c_l do not depend on the direction of wave propagation.

We determine the upper frequency limit from the condition that the total number of oscillations is equal to $3N$. For this, in

Eq. (4.23) we go over to spherical coordinates:

$$\frac{V}{2\pi^2}\sum_{\sigma=1}^{3}\frac{1}{c_\sigma^3}\int_0^{\omega^*}\omega^2\,d\omega=3N \tag{4.28}$$

Substituting c_t and c_l, we obtain

$$\omega^*=\left[\frac{18\pi^2 N}{(2c_t^{-3}+c_l^{-3})\,V}\right]^{1/3} \tag{4.29}$$

Condition (4.28) is such that at high temperatures the law $E = 3N\theta$ is obtained automatically.

At medium temperatures $\theta \sim h\omega^*$. In the integral (4.26), we substitute $k = \omega^*/c_\sigma$ for ∞ as the upper limit. Then the expression for energy has the form

$$E=\frac{Vh}{2\pi^2}\left(\frac{2}{c_t^3}+\frac{1}{c_l^3}\right)\int_0^{\omega^*}\frac{\omega^3\,d\omega}{e^{h\omega/\theta}-1} \tag{4.30}$$

Changing to a new variable $x = h\omega/\theta$ and denoting $h\omega^* = \theta_D$, we can rewrite the expression for lattice energy thus:

$$E=\frac{V}{2\pi^2}\left(\frac{2}{c_t^3}+\frac{1}{c_l^3}\right)\frac{\theta^4}{h^3}\int_0^{\theta_D/\theta}\frac{x^3\,dx}{e^x-1} \tag{4.31}$$

At low temperatures $\theta_D \gg \theta$, therefore the upper limit in the integral is replaced by infinity. The integral is equal to $\pi^4/15$, and for the energy we obtain the expression

$$E=\frac{\pi^2}{30}\frac{V\theta^4}{h^3}\left(\frac{2}{c_t^3}+\frac{1}{c_l^3}\right) \tag{4.32}$$

The exact formula (4.27) assumes the same form if we replace c_σ by c_t and c_l, which are independent of direction.

Going over from energy to specific heat and replacing the coefficient of θ^4 with the help of Eq. (4.29) in terms of θ_D, we obtain

$$C=\frac{12}{5}\pi^4 N\left(\frac{\theta}{\theta_D}\right)^3 \tag{4.33}$$

Comparing this formula with experimental data on specific heat at low temperatures, we can determine θ_D, the *Debye temperature.* But it is also determined directly by calculating c_t and c_l from the elastic properties of bodies. Usually the error in the values of θ_D found by the two methods is of the order of 10 K, which justifies Debye's approximation.[6]

---

[6] Here are some estimates of θ_D for various elements: Pb, 88 K; Na, 172 K; Cu, 315 K; Fe, 453 K; Be, $\approx$1000 K; diamond, 1860 K.

The Mössbauer Effect. Let us examine one more important application of the theory of crystal lattice oscillations. It concerns the mechanism of quantum absorption and emission by the nuclei of a crystal.

When an atomic nucleus emits a quantum, in accordance with the law of conservation of linear momentum, it acquires a momentum $h\omega/c$, where ω is the frequency of the emitted quantum. If the nucleus was at rest at the instant of emission, its kinetic energy after emission is

$$\frac{p^2}{2m} = \frac{(h\omega)^2}{2mc^2} = h\omega\,\frac{h\omega}{2mc^2}$$

where m is the mass of the nucleus.

Let us make evaluations for a typical case. Let $h\omega \sim 10^5$ eV and the atomic weight $A = 100$, so that the rest energy of the nucleus $mc^2 = 9 \times 10^{10}$ eV. Then the kinetic energy of the nucleus's recoil is 0.5×10^{-5} of the energy of the emitted quantum.

An excited nucleus capable of emitting a quantum possesses a known lifetime. The characteristic value of this time Δt for our example is of the order of 10^{-9} s. But a state with a lifetime of Δt possesses an uncertainty in energy of $\Delta E \sim h/\Delta t \sim 10^{-17}$ erg (see [31.37]) or 10^{-5} eV. At the same time the recoil energy of the nucleus is equal to $10^5 \times 0.5 \times 10^{-6} = 0.5 \times 10^{-1}$ eV, which is 5000 times greater than ΔE. This means that an emitted quantum striking an identical nucleus in an unexcited state does not possess sufficient energy to excite it. For a quantum to be capable of exciting the nucleus from which it is emitted, its energy must not deviate from the centre of the emission line by more than 10^{-5} eV (the line width). But owing to recoil it loses 0.5×10^{-1} eV. Similar reasoning is applicable to quantum absorption, when a nucleus receives a linear momentum.

In 1958, Rudolf L. Mössbauer discovered that, if an emitting nucleus is not free and is part of a crystal lattice, there exists a finite probability of such an absorption or emission in which the recoil momentum is transferred not to the individual nucleus but to the crystal as a whole. The recoil energy in this case is smaller by the number of atoms contained in the crystal, for example by a factor of 10^{22}. Then all the effects associated with line width will manifest themselves. In particular, a quantum emitted by a certain nucleus in a certain excited state can be absorbed by similar nonexcited nuclei. The condition for this is that the electromagnetic transition in the nucleus is not responsible for the emission or absorption of a single oscillation quantum of the crystal lattice. It is apparent that in that case only the lattice as a whole can acquire the recoil energy. Thanks to the discrete quality of transitions in quantum systems, there exists a finite probability of such a situation.

The Pound and Rebka Experiment. We shall describe one extremely important application of the Mössbauer effect. Let us consider the work done in lifting two identical nuclei, one in the ground state and the other in an excited state, to a height z. According to the mass-energy relationship, the mass of the excited nucleus is greater than the mass of the unexcited one by $\hbar\omega/c^2$. Hence, the work done in raising it to the height z must be greater by $(\hbar\omega/c^2)gz$ provided exactly the same mass enters the gravitation law as the expression for energy or linear momentum. The energy of the emitted quanta increases by the factor $(1 + gz/c^2)$. For example, at $z = 10^3$ cm the increase in energy or frequency corresponds to the factor $(1 + 10^{-15})$. Can such a quantum excite a nucleus at ground state located 10^3 cm lower?

For the raising of a nucleus to a height z to take the quanta it emits out of resonance with unexcited states of the nucleus, the width of the upper level ΔE must be so small that a change in frequency by a factor of 10^{-15} of its total value would take the quantum outside the natural line width. Obviously, only the Mössbauer effect can assure such precision in measuring the frequency.

In the experiments of R. V. Pound and G. A. Rebka, Jr., the quantum went out of resonance only partly, but quite sufficiently to quantitatively confirm the expected effect.

The frequency of electromagnetic oscillations provides a measure of time. Hence, time passes at a different rate at points with different gravitational potentials. This is a requirement of Einstein's theory of gravitation (the general theory of relativity).

EXERCISES

1. Write the formula for the wavelength distribution of black-body radiation energy. Find the wavelength for which the energy is greatest.

Solution. Proceeding from the fact that $\omega = 2\pi c/\lambda$, we have

$$dE(\lambda) = \frac{16\pi^2 V\hbar c}{\lambda^5 \left[e^{2\pi\hbar c/(\lambda\theta)} - 1\right]} d\lambda$$

The maximum energy per unit interval of λ is determined from the equation $2\pi\hbar c/(\theta\lambda_{max}) = 4.965$.

2. Show that the Bose distribution can be developed by considering the equilibrium between arbitrary bosons and a Boltzmann gas.

Solution. Let the energy of a Boltzmann particle be ε and that of a boson, η. Consider a process in which the transition $\varepsilon + \eta \rightarrow \varepsilon' + \eta'$ occurs, that is, one in which the interaction of particles changes their initial states with energies ε and η to states with energies ε' and η'. In equilibrium the

following balance equation must hold:

$$W_{\varepsilon,\,\eta;\,\varepsilon',\,\eta'} N_\varepsilon n_\eta (1+n_{\eta'}) = W_{\varepsilon',\,\eta';\,\varepsilon,\,\eta} N_{\varepsilon'} n_{\eta'} (1+n_\eta)$$

where $W_{\varepsilon,\,\eta;\,\varepsilon',\,\eta'}$ is the probability of direct transition, and $W_{\varepsilon',\,\eta';\,\varepsilon,\,\eta}$ is that of reverse transition. We have taken account here of what was said of stimulated boson transitions. Assuming for simplicity that $g_{\varepsilon,\,\eta} = g_{\varepsilon',\,\eta'}$ and using the fact that in that case $W_{\varepsilon,\,\eta;\,\varepsilon',\,\eta'} = W_{\varepsilon',\,\eta';\,\varepsilon,\,\eta}$, we obtain

$$n_\eta = \left[\exp\left(\frac{\eta-\mu}{\theta}\right)-1\right]^{-1}, \qquad n_{\eta'} = \left[\exp\left(\frac{\eta'-\mu}{\theta}\right)-1\right]^{-1}$$

if only the Boltzmann particles obey the distribution

$$N_\varepsilon = \exp\left(\frac{\mu-\varepsilon}{\theta}\right), \qquad N_{\varepsilon'} = \exp\left(\frac{\mu-\varepsilon'}{\theta}\right)$$

Stimulated emission leads to the Bose distribution law.

3. For black body radiation, find the total number of quanta,

$$n = \frac{V}{\pi^2 c^3} \int_0^\infty \frac{\omega^2\, d\omega}{e^{\hbar\omega/\theta}-1} = \frac{V\theta^3}{\pi^2 \hbar^3 c^3} \int_0^\infty \frac{x^2\, dx}{e^x - 1}$$

at a given temperature.

Solution. Expand the integrand in a series (see Appendix to Part I)

$$\frac{1}{e^x-1} = \sum_{k=1}^{\infty} e^{-kx}$$

Hence

$$\int_0^\infty \frac{x^2\, dx}{e^x-1} = \sum_{k=1}^{\infty} \int_0^\infty e^{-kx} x^2\, dx = \sum_{k=1}^{\infty} \frac{1}{k^3} \int_0^\infty e^{-y} y^2\, dy = 2 \sum_{k=1}^{\infty} \frac{1}{k^3}$$

The sum is approximately 1.20, and so

$$n \approx \frac{2.4}{\pi^2} \frac{V\theta^3}{\hbar^3 c^3}$$

4. Several atoms are in a linear chain configuration. Denoting the displacement of the sth atom as a_s, and the force acting between the sth and $(s+1)$st atoms as $\alpha(a_{s+1} - a_s)$, where the displacement is along the chain and changes the distances between atoms, write and solve the oscillation equations of the chain. Neglect the interaction between nonadjacent atoms.

Solution. The oscillation equation for the sth atom is

$$m\ddot{a}_s = \alpha\,(a_{s+1} + a_{s-1} - 2a_s)$$

We look for the solution in the form

$$a_s = b\,(t)\, e^{isf}$$

Substituting this into the equation, we obtain, after cancelling out e^{isf},

$$m\ddot{b}(t) = \alpha b(t)(e^{f}+e^{-f}-2) = 2b(t)\,\alpha(\cos f - 1)$$
$$= -4\alpha\, b(t)\sin^2(f/2)$$

Hence, the oscillation frequency for a given value of f is

$$\omega_f = 2\left(\frac{\alpha}{m}\right)^{1/2}\left|\sin\frac{f}{2}\right|$$

If the distance between the atoms is d, then $s = x/d$, where x is the equilibrium position of the sth atom. Introducing the notation $f/d = k$, we obtain

$$e^{ifs} = e^{ikx}$$

Thus, if length is measured in units of d, then f can be regarded as a wave vector. For small values of f frequency is proportional to $|f|$:

$$\omega_f = \sqrt{\frac{\alpha}{d}}\,|f|$$

5. Solve the same problem assuming the even atoms in the chain to be of mass m_1 and the odd atoms of mass m_2.

Answer.

$$\omega_f^2 = \alpha\left\{\frac{m_1+m_2}{m_1m_2} \mp \left[\left(\frac{m_1+m_2}{m_1m_2}\right)^2 - \frac{\sin^2 f}{m_1m_2}\right]^{1/2}\right\}$$

The upper sign corresponds to the acoustic oscillation branch, and the lower to the optical branch.

5

THE BOSE DISTRIBUTION

Choice of Sign of μ. The Bose distribution at low temperatures has some very peculiar properties. We shall assume that the atoms have no spin (as, for example, helium atoms with atomic weight 4). Both the electrons in the shell of a helium atom and the protons and neutrons in the helium nucleus are in the ^1S state. They all come in pairs and their spins, according to Pauli's exclusion principle, are antiparallel. Therefore the resultant spin is zero.

By Eq. (1.30) the weight of a state of a spinless particle is

$$dg(\varepsilon) = \frac{Vm^{3/2}\varepsilon^{1/2}\,d\varepsilon}{2^{1/2}\pi^2\hbar^3} \tag{5.1}$$

The normalization condition has the form

$$\frac{Vm^{3/2}}{2^{1/2}\pi^2 h^3}\int_0^\infty \frac{\varepsilon^{1/2}\,d\varepsilon}{e^{(\varepsilon-\mu)/\theta}-1}=N \tag{5.2}$$

This condition can be satisfied only for negative μ. Indeed, if we assume μ greater than zero, the denominator of the integrand will be negative at $\varepsilon < \mu$ since then $e^{(\varepsilon-\mu)/\theta} < 1$. But this is impossible because the distribution function is by its very nature a positive quantity. Therefore $\mu < 0$. At high temperatures the Bose distribution passes into the Boltzmann distribution according to (2.6)

The Sign of $\partial\mu/\partial\theta$. When the temperature decreases, μ decreases in absolute value. This can be shown in general form with the help of Eq. (5.2). Differentiating this equation as an implicit function, we have

$$\frac{\partial\mu}{\partial\theta}=-\frac{\partial}{\partial\theta}\int_0^\infty \frac{\varepsilon^{1/2}\,d\varepsilon}{e^{(\varepsilon-\mu)/\theta}-1}\Bigg/\frac{\partial}{\partial\mu}\int_0^\infty \frac{\varepsilon^{1/2}\,d\varepsilon}{e^{(\varepsilon-\mu)/\theta}-1}$$

$$=-\int_0^\infty \frac{(\varepsilon-\mu)\,\varepsilon^{1/2}e^{(\varepsilon-\mu)/\theta}\,d\varepsilon}{\theta^2\,(e^{(\varepsilon-\mu)/\theta}-1)^2}\Bigg/\int_0^\infty \frac{\varepsilon^{1/2}e^{(\varepsilon-\mu)/\theta}\,d\varepsilon}{\theta\,(e^{(\varepsilon-\mu)/\theta}-1)^2} \tag{5.3}$$

The integrals in (5.3) are taken with respect to essentially positive quantities ($\varepsilon-\mu > 0$ because $\mu < 0$), and therefore $\partial\mu/\partial\theta < 0$. Hence, when θ decreases, the absolute value of μ diminishes monotonically, since μ must increase.

We shall now show that μ vanishes for a temperature other than zero. For this in (5.2) we put $\mu = 0$ and find the corresponding value $\theta = \theta_0$:

$$\frac{Vm^{3/2}}{2^{1/2}\pi^2 h^3}\int_0^\infty \frac{\varepsilon^{1/2}\,d\varepsilon}{e^{\varepsilon/\theta_0}-1}=\frac{Vm^{3/2}\theta_0^{3/2}}{2^{1/2}\pi^2 h^3}\int_0^\infty \frac{x^{1/2}\,dx}{e^x-1}=N \tag{5.4}$$

The integral represents simply a dimensionless quantity: it is equal to 2.31 (see Appendix to Part I). Therefore Eq. (5.4) holds if θ_0 is not zero.

Bose-Einstein Condensation. What will happen when the temperature is reduced still further? Obviously, μ cannot go over from negative to positive values since this, as pointed out at the beginning of the section, would lead to negative probability values, and μ varies only monotonically if it is at all capable of varying. Therefore, the only possibility is for μ to remain equal to zero, once it has attained its zero value. But then Eq. (5.2) is no longer satisfied if the temperature is less than θ_0 and N remains the same. On the

contrary, it can be seen from (5.4) that if we define the number of particles as

$$N' = \frac{Vm^{3/2}}{2^{1/2}\pi^2 h^3} \int_0^\infty \frac{\varepsilon^{1/2}\, d\varepsilon}{e^{\varepsilon/\theta} - 1} = \frac{2.31 V m^{3/2} \theta^{3/2}}{2^{1/2}\pi^2 h^3} \tag{5.5}$$

for $\theta < \theta_0$ it decreases with the temperature in proportion to $\theta^{3/2}$.

What happens to the remaining $(N - N')$ particles? Unlike light quanta, these particles cannot be absorbed. Therefore, they will pass into a state which is not taken into account in the normalizing integral (5.2). The only state of this kind possesses zero energy: due to the factor $\varepsilon^{1/2}$ it contributes nothing to the integral (5.4). In normalization we can isolate the particles occurring in the zeroth state in a separate term. If a finite number of particles pass into the zero-energy state, they will naturally drop out of the integral. Hence, N particles remain continuously distributed, but for $\mu = 0$. Thus, at temperatures $\theta < \theta_0$ the whole distribution consists of an infinitely narrow "peak" at $\varepsilon = 0$ and particles distributed according to the $(e^{\varepsilon/\theta} - 1)^{-1}$ law. At absolute zero all the particles are in the zeroth state; this state of a Bose gas is, obviously, defined uniquely. A Boltzmann gas would behave in an entirely different way if the temperature tended to zero. All particles would remain within the $e^{(\mu-\varepsilon)/\theta}$ distribution, no matter how small the values of θ.

Liquid Helium. As already stated, since the nuclear and the electron-shell spin of helium with atomic weight 4 is zero, it is subject to Bose statistics. It is therefore of interest to see whether anything like the Bose-Einstein condensation occurs.

At low temperatures helium is liquid and the Bose distribution, which holds for ideal gas, does not apply. S. T. Belyaev showed that strongly interacting Bose particles can also assemble at the zeroth state. Therefore the qualitative aspect of the result obtained for a gas holds. The behaviour of liquid helium can be compared with what the elementary theory of a Bose gas set forth here yields.

Liquid helium does, in fact, undergo a peculiar change of state at temperature 2.19 K (at atmospheric pressure). In a monatomic liquid (and liquid helium is monatomic) it is hard to imagine any spatial rearrangement of atoms. Assuming that, as in gaseous helium, changes in liquid helium are due to a redistribution of particles in momentum space, it is reasonable to compare the actual temperature of liquefaction of helium with the temperature at which Bose-Einstein condensation would have commenced in gaseous helium of the same density.

The density of liquid helium is 0.12 g-cm³. Consequently

$$\frac{N_A}{V} = \frac{0.12}{4} \times 6 \times 10^{23} = 0.18 \times 10^{23}\ \text{cm}^{-3}$$

Hence, by (5.4) the temperature θ_0 is

$$\theta_0=\left(\frac{1.18\times10^{23}\times9.86\times1.41\times1\times18\times10^{-81}}{2.31\times17\cdot1\times10^{-36}}\right)^{2/3}$$
$$=3.86\times10^{-16}\text{ erg}$$

or

$$T_0=2.8\text{ K}$$

which is close to the transition temperature. Note that in the case of a Bose gas the specific heat at the transition point is continuous; only its derivative with respect to temperature experiences a discontinuity. In the case of liquid helium the specific heat has a discontinuity. This shows that the properties of helium as a liquid are important for an understanding of the nature of the transition.

Superfluidity. P. L. Kapitza discovered that below the transition temperature liquid helium possesses a remarkable property: it is capable of passing through the narrowest capillaries without exhibiting any signs of viscosity. This property has been called *superfluidity* (see Sec. 19).

The question of the relation between superfluidity and Bose-Einstein condensation has not yet been fully resolved. The fact that the helium isotope of atomic weight 3 does not exhibit superfluidity[7] speaks in favour of such a relation. The nuclear spin of He^3 is 1/2, so that its atoms are subject to Fermi, and not Bose, statistics. Accordingly, they cannot all pass into the zero state simultaneously. Pauli's exclusion principle prohibits this.

N. N. Bogoliubov showed that a gas closely approximating an ideal gas and consisting of Bose particles possesses the same energy spectrum which, according to Landau's theory, a superfluid should have in low-energy states. Analogous results for strongly interacting particles were obtained by S. T. Belyaev. However, no one has so far succeeded in fully proving that it is precisely liquid helium that should possess superfluidity below the transition point.

EXERCISE

Calculate the energy and pressure of a Bose gas below the transition point.

Solution. The energy of a Bose gas is determined as follows:

$$E=\frac{Vm^{3/2}}{2^{1/2}\pi^2h^3}\,\theta^{5/2}\int_0^\infty\frac{x^{3/2}\,dx}{e^x-1}=\frac{1.78Vm^{3/2}\theta^{5/2}}{2^{1/2}\pi^2h^3}$$

---

[7] At superlow temperatures ($\sim10^{-3}$ K) He^3 passes to the superfluid state owing to the fact that its atoms form bonded pairs. This phenomenon is analogous to superconductivity (see Sec. 43)

(see Appendix to Part I). The pressure is determined from the relationship (2.22):

$$p=\frac{2}{3}\frac{E}{V}=\frac{1.18m^{3/2}\theta^{5/2}}{2^{1/2}\pi^2h^3}$$

Thus, the pressure of a Bose gas below the transition point does not depend on volume and is a function only of temperature. If such a Bose gas is compressed, its particles will pass into the zero-energy state. Conversely, in expansion the particles of a Bose gas will escape out of the zero-energy state until none remain in the state. As expansion continues, the pressure begins to drop. Note that the pressure of black body radiation also depends only on temperature. In compression, a portion of the quanta are simply absorbed and thus taken out of the game. In a Bose gas, which consists of nondestructable particles, instead of absorption transition to zero state takes place.

6

THE FERMI DISTRIBUTION

The Form of the Fermi Distribution Curve and Its Interpretation. The criterion for the transition from quantum to classical statistics is

$$\frac{N}{V}\ll\frac{g^{(0)}}{h^3}\left(\frac{m\theta}{2\pi}\right)^{3/2}$$

according to (2.7). If the inequality is reversed, the statistical distribution displays essentially quantum properties. In this section we shall examine the properties of the Fermi distribution when the inequality

$$\frac{N}{V}\gg\frac{g^{(0)}}{h^3}\left(\frac{m\theta}{2\pi}\right)^{3/2} \tag{6.1}$$

or the equivalent inequality

$$\frac{\mu}{\theta}\gg 1 \tag{6.2}$$

holds.

By Eqs. (1.26) and (1.30) the Fermi distribution has the following form:

$$dn(\varepsilon)=\frac{V(2m^3)^{1/2}\varepsilon^{1/2}\,d\varepsilon}{\pi^2h^3}\left(e^{(\varepsilon-\mu)/\theta}+1\right)^{-1} \tag{6.3}$$

The weight factor 2 has been introduced here on the assumption that $j=1/2$. The first factor in (6.3) represents the total number

of states between ε and $\varepsilon + d\varepsilon$, and the second factor expresses the probability of one of these states, arbitrarily chosen, being occupied. In other words, the Fermi factor characterizes the relative density of occupied states with an energy ε. The function

$$f(\varepsilon) = (e^{(\varepsilon-\mu)/\theta} + 1)^{-1} \tag{6.4}$$

can be interpreted both as a probability and as the mean number of particles per state, taking into account that $f(\varepsilon)$ is always contained between zero and unity. A similar function in the Bose distribution could denote only the mean number of particles in one state

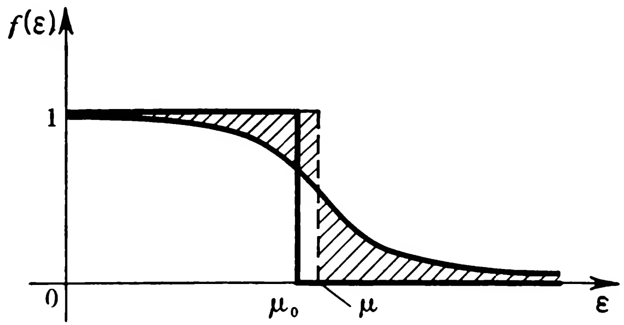


Figure 5

with a given energy, because the function $(e^{(\varepsilon-\mu)/\theta} - 1)^{-1}$ can be greater than unity and must not be interpreted as probability.

Let us see how the curve $f(\varepsilon)$ behaves when $\mu/\theta \gg 1$. At $\varepsilon = 0$ we have

$$f(0) = (e^{-\mu/\theta} + 1)^{-1} \approx 1$$

because from the inequality (6.2) $e^{-\mu/\theta}$ is a small number. As long as ε remains smaller than μ, the quantity $e^{(\varepsilon-\mu)/\theta}$ is also small and $f(\varepsilon)$, like $f(0)$, is close to unity. Only when $\varepsilon - \mu$ is comparable with θ is $e^{(\varepsilon-\mu)/\theta}$ of the order of unity, and $f(\varepsilon)$ begins to decrease perceptibly as ε increases. At $\varepsilon = \mu$ the value of $f(\varepsilon)$ decreases to 1/2:

$$f(\mu) = \frac{1}{e^0 + 1} = \frac{1}{2}$$

For still greater values of ε the function $f(\varepsilon)$ decreases exponentially, because unity can be neglected in comparison with the exponential function. The function $f(\varepsilon)$ becomes the Boltzmann distribution:

$$f(\varepsilon) \sim e^{(\mu-\varepsilon)/\theta}$$

The Bose distribution has the same limiting form. The curve $f(\varepsilon)$ is presented approximately in Figure 5. The region of values of

ε at which $f(\varepsilon)$ varies from unity to zero has a width of the order of θ since the $e^{(\varepsilon-\mu)/\theta}$ is comparable with unity only if $\varepsilon - \mu \sim \theta$: for smaller ε the exponent is considerably smaller than unity, and at larger ε it is considerably greater.

The Fermi Distribution at Absolute Zero. We shall call the region of f from unity to zero the *spread* of the Fermi distribution. As the temperature decreases, the spread narrows, and at absolute zero becomes a sharp discontinuity, so that the distribution function has the form of a step. We used this form of f in [Sec. 33] in developing the Thomas-Fermi potential in an atom. In Figure 5, the step distribution is shown by the solid line. The value of μ at absolute zero is denoted μ_0. Consequently, at $\theta = 0$ all states with energy less than μ_0 are occupied with unit probability, that is, with certainty, and states with energy greater than μ_0 are empty (with certainty).

The Criterion of the Proximity of the Fermi Distribution to its Form at Absolute Zero. The Fermi distribution can be imagined in a momentum space. If momentum is defined as $p_0^2/(2m) = \varepsilon_0 = \mu_0$, it will be a boundary. All states with $\varepsilon \leqslant \varepsilon_0$ are occupied and all states with $\varepsilon > \varepsilon_0$ are empty. The surface $\varepsilon = \varepsilon_0$ is called the *Fermi surface*.

Let us repeat briefly the evaluation of the quantity $\varepsilon_0 = \mu_0$ done in [Sec. 33]. Taking account of what was said of the step character of the function $f(\varepsilon)$, we make use of the distribution (6.3), which yields

$$N = \int_0^{\varepsilon_0} dn(\varepsilon) = \frac{V(2m^3)^{1/2}}{\pi^2 h^3} \int_0^{\varepsilon_0} \varepsilon^{1/2}\, d\varepsilon = \frac{V(2m)^{3/2}\varepsilon_0^{3/2}}{3\pi^2 h^3} \tag{6.5}$$

whence

$$\varepsilon_0 = \mu_0 = 3^{2/3}\pi^{4/3}\,\frac{h^2}{2m}\left(\frac{N}{V}\right)^{2/3} \tag{6.6}$$

The state of a Fermi gas at absolute zero as a whole is determined, as in general in quantum statistics, by the states occupied by individual particles, not by what particles occupy them. In the present case all the states within the surface of the limiting-energy sphere $\varepsilon = \varepsilon_0$ are occupied by particles.

At temperatures approaching absolute zero thermal excitation can be imparted only to those particles whose energy approximates $\varepsilon = \varepsilon_0$. Indeed, as long as $\varepsilon < \varepsilon_0$, thermal excitation of the order of θ cannot be transmitted to a particle whose state corresponds to an energy lying deep inside the Fermi surface $\varepsilon = \varepsilon_0$, since all the states between $\varepsilon < \varepsilon_0$ and the Fermi surface $\varepsilon = \varepsilon_0$ are occupied and the energy θ is insufficient to eject a particle outside the Fermi surface. Only particles whose energy differs from ε_0 by an amount

of the order of θ can occupy empty places. Deeper states at such a temperature are densely filled. Hence, the occupation probability is almost unity at all energies $\varepsilon < \varepsilon_0$ and decreases to zero in a region (close to ε_0) whose width is of the order of θ, as shown in Figure 5.

The criterion that the curve form approaches a step function is, as is apparent from what has been said, the relationship

$$\theta \ll \varepsilon_0$$

which agrees with (6.1) within a numerical factor. We shall now show that the low-temperature-approximation criterion (6.1) differs greatly from the conventional.

Conduction electrons in metals are usually treated as an ideal gas, neglecting the action of lattice ions and their interactions, which is essential for application of the gas-state concept. Ion action breaks down the simple relationship

$$\varepsilon = \frac{p^2}{2m}$$

between energy and momentum. For that reason in most metals the Fermi surface in no way resembles a sphere. The main exception is alkali metals. Taking into account the complex dependence of energy on momentum, we could still treat the electrons as a gas in a field of lattice ions, which is in fact sometimes done. So far it has proved impossible strictly to take into account the interactions of electrons.

Nevertheless, the concept of electrons in a metal as a Fermi gas of noninteracting particles usually produces results in good agreement with experimental data. This fully justifies employing the concept for a simple evaluation. As an example, let us take metallic sodium, whose last electron is weakly bound with the atom and easily separates from it in the lattice. Accordingly, in sodium, as in alkali metals in general, the dependence $\varepsilon = p^2/(2m)$ holds, and the Fermi surface is approximated by a sphere.

The density of metallic sodium is 0.97, its atomic weight is 23, therefore unit volume contains

$$\frac{0.97}{23} \times 6.02 \times 10^{23} = 0.25 \times 10^{23}$$

atoms and as many conduction electrons. Hence, from (6.6)

$$\varepsilon_0 = 4.1 \times 4.6 \frac{1.12 \times 10^{-54}}{1.8 \times 10^{-27}} 0.8 \times 10^{-17} = 4.8 \times 10^{-12}$$

which corresponds to 34 800 K. The numbers here are in the same sequence as they occur in (6.6).

Consequently, at all temperatures at which sodium can be treated as a metal, the electron gas in it approximates a Fermi gas at abso-

lute zero. Similar results are also obtained for nonalkali metals, though with a less reliable value of electron density.

At equal gas density the Fermi energy of electrons is 1840 times greater than the Fermi energy of protons. The rest energy of a separate neutron is somewhat greater than that of a proton. That is why a proton in a free state is more stable than a neutron, which decays spontaneously into a proton, electron, and antineutrino[8]. However, in a superdense Fermi gas it is more advantageous for matter to pass into a neutron state, in which there are no electrons with high Fermi energy and the total energy of the neutron Fermi gas is less than the energy of a Fermi gas comprising electrons and protons. There is substantial evidence that neutron stars actually do exist.

Compressibility of Alkali Metals. Let us derive a formula for the compressibility of a Fermi gas at absolute zero. From (6.5) and (6.6), the energy at absolute zero is

$$E = \int_0^{\varepsilon_0} \varepsilon \, dg(\varepsilon) = \frac{V (2m)^{3/2} \varepsilon_0^{5/2}}{5\pi^2 h^3} \tag{6.7}$$

From the Bernoulli formula (2.22), the pressure equals two-thirds the energy density, that is

$$p = \frac{2}{15} \frac{(2m)^{3/2} \varepsilon_0^{5/2}}{\pi^2 h^3} = \frac{3^{2/3} \pi^{4/3}}{5} \frac{h^2}{m} \left(\frac{N}{V}\right)^{5/3} \tag{6.8}$$

whence

$$-\frac{\partial}{\partial p} \ln V = \frac{3}{5p} = \frac{3^{1/3}}{\pi^{4/3}} \frac{m}{h^2} \left(\frac{N}{V}\right)^{-5/3}$$

$$= 0.273 \times 10^{27} \left(\frac{N}{V}\right)^{-5/3} \text{ bar}^{-1} \tag{6.9}$$

Ya. I. Frenkel noted that the compressibility of alkali metals approximates that of an electron gas. Indeed, expressing N/V in terms of atomic weight and density, we obtain the following table:

	Li	Na	K	Rb	Cs
$-\frac{\partial \ln V}{\partial p} \times 10^6$ from Eq. (6.9)	4.7	13	37	52	79
$-\frac{\partial \ln V}{\partial p} \times 10^6$ from experimental data	8	15	32	40	61

In a real crystal lattice there are, of course, not only forces of repulsion between particles but also cohesive forces. The equilib-

[8] It is convenient to treat the neutrino emitted in such a decay process as an antiparticle, thereby preserving the difference between the number of particles and antiparticles in the decay process.

rium between the forces of cohesion and repulsion determines the characteristic volume which every condensed body (solid or liquid) has in the absence of external pressure. Normal atmospheric pressure provides a force that is vanishingly small in comparison with the tremendous forces that keep bodies within their volumes. In order to change the volume of a body by only one per cent pressures of the order of tens of thousands of atmospheres are required.

The agreement of theoretical and experimental data presented here indicates that when alkali metals are compressed, the cohesive forces change insignificantly in comparison with the forces of repulsion. It is further conceivable that the state of the valence electrons in alkali metals is perturbed to a comparatively small degree by atomic cores and to some extent approximates an electron gas. Compression has little effect on the extremely dense electron shells of the atomic cores. Therefore the compressibility of alkali metals approximates that of an ideal Fermi gas. That this should be so is, of course, not at all obvious beforehand.

Specific Heat of a Fermi Gas. In conclusion we shall consider a Fermi gas not at absolute zero but at temperatures which, though differing from zero, satisfy the inequality (6.1).

To begin with, it is convenient to derive a general evaluation of the integral entering the Fermi distribution for $\theta \ll \varepsilon_0$. Consider the integral

$$I=\int_0^\infty \frac{d\gamma(\varepsilon)}{d\varepsilon}\,\frac{1}{e^{(\varepsilon-\mu)/\theta}+1}\,d\varepsilon \tag{6.10}$$

where $\gamma(\varepsilon)$ is a power function, for example, $\varepsilon^{1/2}$, $\varepsilon^{3/2}$, etc.

The Fermi distribution curve (Fig. 5) coincides with itself in a rotation through 180° with respect to the point $f = 1/2$, $\varepsilon = \mu$ if we ignore the region where $\varepsilon < 0$, which makes an exponentially small contribution to the integral at $\mu \gg \theta$. Hence, if we approximate $\gamma(\varepsilon)$ by a series expansion

$$\gamma(\varepsilon)=\gamma(\mu)+(\varepsilon-\mu)\,\gamma'(\mu)+\frac{1}{2}(\varepsilon-\mu)^2\,\gamma''(\mu)+\dots \tag{6.11}$$

the integral with the odd functions (with respect to $\varepsilon = \mu$) vanishes. After integrating by parts the integral of the even term $(\varepsilon - \mu)^2$ gives a contribution proportional to θ^2, which is evident if we replace $(\varepsilon - \mu)/\theta$ by a dimensionless quantity x.

To calculate the specific heat of a Fermi gas, we must find the normalization integral (at $\gamma'(\varepsilon) = \varepsilon^{1/2}$) and the mean energy (at $\gamma'(\varepsilon) = \varepsilon^{3/2}$). At a temperature differing from absolute zero both receive terms proportional to θ^2. Hence, the increment to specific heat, that is, $\partial\varepsilon/\partial\theta$, will be linear in θ. For example, an evaluation for sodium yields $\varepsilon_0 = 34\,800$ K, so that at room temperature $\theta/\varepsilon_0 \sim$

~ 0.01. The specific heat of a Fermi gas per electron at room temperature is equal to 0.05. It must be compared with the specific heat of a Boltzmann gas, which, according to Section 2, is equal to 1.5 (if θ is measured in ergs, the specific heat C is a dimensionless quantity).

It will be readily appreciated why the specific heat of a Fermi gas is considerably less than that of a Boltzmann gas: for a Fermi distribution not all electrons are capable of responding to thermal excitation, only those whose energy approaches the limiting energy. That is why the specific heat of a Fermi gas is equal to only several percentage points of N. The specific heat $3N/2$ is obtained only when all electrons are capable of responding to thermal excitation.

The old quantum electron theory of metals encountered considerable difficulties in that at room temperature the electron gas in a metal has no experimentally appreciable specific heat. The specific heat of metal does not exceed 3 per atom (see Eq. (4.24)). Yet if the number of electrons equals the number of atoms, a metal would, by classical statistics, have a specific heat of $3 + 3/2 = 9/2$ per atom, which is never observed. The significance of Fermi statistics for electrons in a metal was shown by Arnold J. W. Sommerfeld, who developed the formula for the specific heat of a Fermi gas.

At low temperatures the specific heat of a metal crystal lattice is proportional to θ^3 (see Eq. (4.27)). Hence, if the temperature is low enough, the electron specific heat begins to predominate and can be measured. Experiments show that at very low temperatures the specific heat of metals actually is proportional to θ. It can be seen from (4.33) that, knowing specific heat, we can determine the number of electrons per atom. It is interesting to note that bismuth (in many ways an atypical metal) has very few conduction electrons.

EXERCISES

1. Find the equilibrium concentration of electrons in a volume devoid of charges, that is, of matter, at low temperatures.

Solution. Instead of the conservation of the number of particles, we must take into account the conservation of charge in the creation and annihilation of electron-positron pairs [Sec. 37]. Denoting the number of electrons in a given quantum state as f^-, and the number of positrons as f^+, we have, instead of (1.17), the following supplementary condition:

$$\sum_k g_k (f_k^- - f_k^+) = 0$$

Determining the f^- and f^+ yielding the maximum of the function P with the given supplementary condition, we obtain the distribution functions

for electrons and positrons:

$$f^- = (e^{(\varepsilon-\mu)/\theta}+1)^{-1}, \qquad f^+ = (e^{(\varepsilon+\mu)/\theta}+1)^{-1}$$

Under the conditions of the problem the total number of electrons must equal the total number of positrons:

$$\int_0^\infty \frac{\varepsilon^{1/2}\,d\varepsilon}{e^{(\varepsilon-\mu)/\theta}+1} = \int_0^\infty \frac{\varepsilon^{1/2}\,d\varepsilon}{e^{(\varepsilon+\mu)/\theta}+1}$$

This equation has only one solution with respect to μ: $\mu = 0$. Hence, the total number of electrons in unit volume is

$$\frac{2\times 4\pi}{(2\pi h)^3}\int_0^\infty \frac{p^2\,dp}{e^{\varepsilon/\theta}+1}$$

Calculate this integral for $\theta \ll mc^2$. We can take the nonrelativistic approximation $\varepsilon = mc^2 + p^2/(2m)$ for the energy, and the distribution function according to Boltzmann as $e^{-\varepsilon/\theta}$. From this we obtain the expression for the equilibrium concentration of electrons:

$$\frac{1}{\pi^2 h^3}\, e^{-mc^2/\theta}\int_0^\infty \exp\left(-\frac{p^2}{2m\theta}\right) p^2\,dp$$

$$= \frac{1}{2^{1/2}\pi^{3/2}}\left(\frac{mc}{h}\right)^3\left(\frac{\theta}{mc^2}\right)^{3/2} e^{-mc^2/\theta}$$

This quantity becomes equal to 1 cm^{-3} at $\theta = mc^2/64 = 8$ keV. The energy of the electromagnetic field per unit volume at the same temperature is equal to 0.6×10^{18} erg, whereas the available rest energy of an electron *and* positron is only 1.6×10^{-6} erg. The energy of electrons and positrons approximates the electromagnetic-field energy when $\theta \sim mc^2$. If $\theta \gg mc^2$, the rest energy is of no consequence, and $\varepsilon = cp$ for quanta, electrons, and positrons. The electromagnetic field accounts for 1/3 the energy.

2. Find the limiting energy of a superdense electron gas for which the energy-momentum dependence is in the main ultrarelativistic: $\varepsilon = cp$. Determine the density at which the gas may be regarded as ultrarelativistic.

Solution. Instead of Eq. (6.1) we have

$$\frac{4\pi}{3}\frac{\varepsilon^3}{c^3} = \frac{N}{2}\frac{(2\pi h)^3}{V}$$

(see the derivation of [33.26]). Hence

$$\varepsilon_0 = \left(\frac{3}{8\pi}\right)^{1/3}\left(\frac{N}{V}\right)^{1/3} 2\pi hc$$

If

$$\varepsilon_0 \gg mc^2$$

the rest energy can be neglected and the density concondition can be written in the form

$$\frac{N}{V} \gg \frac{1}{3\pi^2} \left(\frac{mc}{h}\right)^3 \approx 10^{30} \text{ electrons/cm}^3$$

(since ε_0 involves $(N/V)^{1/3}$, the inequality must be strong).

The energy of an ultrarelativistic gas is given by the expression

$$\frac{V}{\pi^2 h^2 c^3} \int_0^\infty \frac{\varepsilon^3 \, d\varepsilon}{e^{(\varepsilon-\mu)/\theta}+1}$$

3. Determine the number of electrons ejected by thermal excitation from the surface of a metal in unit time.

Solution. Only those electrons can eject from the metal surface whose velocity component normal to the wall is greater than a certain quantity v_{0x} that satisfies the inequality

$$\frac{m}{2} v_{0x}^2 - \mu(0) \gg \theta$$

In other words, the energy of the ejected electrons differs from the limiting energy by considerably more than θ.

We treat the metal as a potential well of finite depth $mv_{0x}^2/2$. Since the tangential velocity component is conserved in the crossing of the surface, the work function from the bottom of the well is equal to $mv_{0x}^2/2$. The work function from the Fermi boundary μ is equal to $(mv_{0x}^2/2) - \mu$. At temperatures other than zero there are always electrons with energies greater than $mv_{0x}^2/2$. It is mainly these electrons that are emitted from the metal (thermionic emission).

The number of electrons with velocity v_x falling onto one square centimetre of the surface per second is

$$v_x \, dn\,(v_x)$$

where $dn\,(v_x)$ is the density of the electrons having a given velocity projection v_x. Let us write the expression for the density of the electrons whose velocity components lie within a given interval (it is analogous to the expression (6.3)):

$$dn\,(v_x, v_y, v_z) = \frac{2m^3 \, dv_x \, dv_y \, dv_z}{(2\pi h)^3} \left(e^{(\varepsilon-\mu)/\theta}+1\right)^{-1}$$

where $\varepsilon = m\,(v_x^2+v_y^2+v_z^2)/2$. Only those electrons cross the surface of the metal for which the difference $\varepsilon - \mu$ is considerably greater than θ. Therefore, we can go over from a Fermi to a Boltzmann distribution with the same value of μ as in the Fermi distribution. Hence, the required electron flux, calculated according to the "tail" of the Fermi curve, where $\varepsilon - \mu \gg \theta$,

is

$$\frac{2m^3}{(2\pi h)^3} e^{\mu/\theta} \int_{v_{0x}}^{\infty} v_x \exp\left(-\frac{mv_x^2}{2\theta}\right) dv_x$$

$$\times \int_{-\infty}^{\infty} \exp\left(-\frac{mv_y^2}{2\theta}\right) dv_y \int_{-\infty}^{\infty} \exp\left(-\frac{mv_z^2}{2\theta}\right) dv_z$$

$$= \frac{2m^3}{(2\pi h)^3} e^{\mu/\theta} \frac{\theta}{m} \exp\left(-\frac{mv_{0x}^2}{2\theta}\right) \times \frac{2\pi\theta}{m}$$

$$= \frac{m\theta^2}{2\pi^2 h^3} \exp\left[\frac{1}{\theta}\left(\mu - \frac{mv_{0x}^2}{2}\right)\right]$$

If an electric field is applied to the metal, the maximum current output at a given temperature (the saturation current) is determined by this formula. Since it relates to electrons in a metal, the quantity μ is close to μ_0, that is, the limiting energy at absolute zero, and does not depend upon temperature.

Note that if a very strong electric field is applied to the metal, the emitted electrons will have overcome the potential barrier appearing under such conditions at the boundary (field emission). But this requires very strong fields. Field emission is analogous to the ionization of atoms in the Stark effect (see [Sec. 33]).

7

GIBBS STATISTICS

In this section we shall consider the bases of the general statistical method of Josiah W. Gibbs, which is applicable to all systems comprising a sufficiently large number of particles, irrespective of whether these systems are liquid, solid, or gaseous. It is very difficult to substantiate this method on the basis of the concept of an arbitrary system of interacting particles. Realizing this, Gibbs did not attempt to derive his results from the equations of mechanics: he simply proceeded from the general distribution function enunciated by him. His works and all the subsequent developments of physics have shown that his method is in fact universal.

When quantum statistics appeared, it too fitted beautifully into Gibbs' general propositions. Their derivation from the equations of quantum mechanics is also extremely difficult, but from the

physical point of view this is nevertheless somewhat simpler than from the equations of classical mechanics, in which there is no place for probability. The tremendous efforts of many outstanding mathematicians to offer a classical substantiation of Gibbs statistics are mainly of methodological interest.

But as a matter of principle it would be well to substantiate the statistical method in such a way that its agreement with experience would follow from agreement of experience with the laws of mechanics (both classical and quantum). In the absence of such strict proof we must be satisfied with semi-intuitive considerations that support at least the natural character of Gibbs' statistical approach.

Quasi-Closed Systems. As was shown in Section 1, statistics deals with systems in weak interaction with the surroundings. We shall call such systems *quasi-closed.* The interaction does not essentially disturb the structure of the quasi-closed system, but it makes for transitions between those of its states which correspond to close separate energy levels of a closed system. In a system consisting of a sufficiently large number of particles, an energy interval ΔE due to the quasi-closed nature of the system contains a vast number of separate energy levels (or, more precisely, states) corresponding to separate, very close energy levels of an ideally closed system. It is this that makes the application of statistics possible.

Statistical Equilibrium. As was shown in Section 1, all these separate states are equiprobable: the proof did not require the system to be necessarily gaseous. In other words, the system spends the same time in every state. In studying the behaviour of a macroscopic system it is important to know not its detailed state (characterized by a certain wave function) but a large group of states, to which the detailed state of the system belongs most of the time.

It was such groups of states that were considered as most probable in the preceding sections. We found that a gas was subject to the Bose or Fermi distributions, depending on whether its particles had integral or half-integral spin. If the particles of a gas approximate the most probable distribution at constant conditions of interaction with the external medium, the state of the gas will all the time approximate the most probable. The possibility of a more or less substantial deviation from the most probable state is vanishingly small.

The totality of equiprobable microscopic states in which a system spends most of the time is called the *state of statistical equilibrium* of the system. In Section 4 it was called thermal equilibrium. It will be shown further on that in the cases considered these two equilibrium concepts are equivalent. The statistical equilibrium state is defined in far less detail than is done for states in quantum mechan-

ics, but sufficiently for a macroscopic description of the system as a whole.

The concept of statistical equilibrium can be applied to any sufficiently large system of particles irrespective of whether they interact weakly, like the particles of an ideal gas, or strongly, as in liquid and solid states. It will be recalled that the proof of the equiprobability of microstates in Section 1 did not assume that the system consisted of noninteracting particles. The state of any system in the macroscopic sense is the more probable the greater the number of microstates it includes. Here, only those microstates are taken into account which are compatible with the energy conservation law, that is, belong to the energy interval ΔE of a quasi-closed system.

Probability Distribution in Subsystems. Instead of considering a quasi-closed system in an external medium it is more convenient to assume a large perfectly closed system and divide it into separate quasi-closed subsystems each of which is of macroscopic dimensions, that is, consists of a vast number of molecules. Within it we can distinguish the internal (bulk) part and the surface layer where it borders on other subsystems. A subsystem is quasi-closed if the surface layer across which it interacts with neighbouring subsystems does not appreciably affect processes taking place within the volume. Interactions within a subsystem lead to the establishment of equilibrium within it, while interactions between subsystems lead to the statistical equilibrium of the system as a whole.

Suppose that equilibrium has been established within a subsystem. What is the probability that its energy lies within the interval from E to $E + dE$? To this interval there correspond $g(E)$ equiprobable microstates. In Section 1 the function $g(E)$ was called the weight of a state with the energy E.

Since all the separate microstates are equiprobable, the probability that the subsystem is in a state with the energy in the interval dE is proportional to $g(E)$:

$$P(E) = \rho(E)\, g(E) \tag{7.1}$$

where $\rho(E)$ is a function we have to determine.

Separate quasi-closed subsystems can be treated as very large molecules of a Boltzmann gas. Since the macroscopic subsystems are distinguishable from one another, it is natural to assume that such a "gas" is subject to Boltzmann statistics. We can, therefore, expect the distribution function to be of the Boltzmann form:

$$\rho(E) = e^{-E/\theta}$$

This conclusion is of a tentative, intuitive nature. A more strict derivation based on properties of the function $\rho(E)$, which will now be established, is presented further on.

Liouville's Theorem. We shall prove that the function $\rho(E)$ is constant during the time interval in which a quasi-closed system may be regarded as closed (that is, the other subsystems do not appreciably affect its state).

The weight of a state, $g(E)$, is determined by the number of microstates whose energies lie within the interval from E to $E + dE$. Each of these microstates is characterized by a definite set of integrals of motion. In an ideal gas this may be the totality of the momenta of separate molecules, their vibration and quantum rotation numbers. In the most general case it is the number of states of the quantum system, each of which is characterized by a definite wave function. In other words, it is a quite definite mechanical description of the system with regard to its energy interval dE. Hence, as long as a quasi-closed system can be treated as strictly closed, that is, over a brief but finite period of time, $g(E)$ is a constant quantity.

The probability $P(E)$ is defined as $\lim t(E)/t$ where t tends to infinity (see Sec. 1). Here, t is the time during which the total closed system, which includes the given quasi-closed subsystem, is observed. Therefore, by the very meaning of $P(E)$ as a resultant (mean) function for time intervals of any duration it cannot depend on time. But if $P(E)$ is a constant quantity and $g(E)$, as a function of the integrals of motion, is also constant, then, from Eq. (7.1), the required function $\rho(E)$ does not depend on time either and is itself an integral of motion. And since all integrals of motion are, at least in principle, known from mechanics, $\rho(E)$ must be a function of them. In other words, $\rho(E)$ cannot depend on quantities that vary with time and, apart from E, depends only on other integrals of motion. More exactly, $\rho(E)$ remains finitely constant not all the time but only within the time intervals when the quasi-closed system can be regarded as closed. The statement concerning the constancy of $\rho(E)$ is known as *Liouville's theorem*. The proof presented here is, in quantum terms, for discrete states. We shall not need Liouville's own classical formulation.

The Theorem of Multiplication of Probabilities. Over a certain time interval quasi-closed subsystems can be treated as independent, and the well-known theorem of the multiplication of probabilities can be applied: the probability of one subsystem occurring in state A and another in state B is equal to the product of the probability of state A and the probability of state B:

$$P_{AB} = P_A P_B \tag{7.2}$$

This assertion is very simply proved with the help of the definition of probability accepted here. If t_A is the time spent by the first subsystem in state A, t_B is the time spent by the second subsystem in state B, and t_{AB} is the time spent by the second subsystem

in state B while the first is in state A, then from the definition of probability we have

$$P_{AB} = \lim_{t\to\infty} \frac{t_{AB}}{t}$$

But this limit can also be represented as follows:

$$P_{AB} = \lim_{t\to\infty} \frac{t_{AB}}{t_A} \times \frac{t_A}{t} = \lim_{t\to\infty} \frac{t_{AB}}{t_A} \times \lim_{t\to\infty} \frac{t_A}{t}$$

$$= \lim_{t\to\infty} \frac{t_B}{t} \times \lim_{t\to\infty} \frac{t_A}{t}$$

We assume that if the systems are independent, the equation

$$\lim_{t_A\to\infty} \frac{t_{AB}}{t_A} = \lim_{t\to\infty} \frac{t_B}{t}$$

must hold, since it does not matter whether we observe the second subsystem continuously or only when the first subsystem is in state A. From this follows Eq. (7.2).

A statement analogous to (7.2) with regard to statistical weights is obvious from their definition, because they refer to different subsystems:

$$g_{AB} = g_A g_B \tag{7.3}$$

Thus, the equations

$$P_{AB} = P_A P_B = g_{AB}\rho_{AB} = g_A\rho_A \times g_B\rho_B$$

must hold.

It follows from Eqs. (7.2) and (7.3) that

$$\rho_{AB} = \rho_A \rho_B \tag{7.4}$$

In other words, the function ρ, which we call the *probability density*, for two independent subsystems is equal to the product of the probability densities for each subsystem. Such a function is called *multiplicative*.

The Gibbs Distribution. It can now be shown that $\rho(E)$ has the expected exponential form. The logarithm of the probability density is an *additive* quantity, that is, it equals the sum of the logarithms of this quantity for each subsystem separately:

$$\ln \rho_{AB} = \ln \rho_A + \ln \rho_B \tag{7.5}$$

In addition, it follows from Liouville's theorem that $\ln \rho$ is an integral of the motion. Hence, this integral of the motion is additive.

In [Sec. 4] it was shown that there exist the following additive integrals of the motion for a closed system: energy, linear momentum,

and angular momentum. For $\ln \rho$ to be an additive integral of motion it must be *linearly* dependent on energy, linear momentum, and angular momentum. If we select a frame of reference in which the system as a whole does not move or rotate, the linear and angular momenta are zero, and only the linear dependence upon energy

$$\ln \rho = aE + b \tag{7.6}$$

remains. The factor a must be the same for all subsystems of the larger system since otherwise $\ln \rho$ would not possess the properties of an additive function. If a is the same for two subsystems, we obtain for these subsystems

$$\begin{aligned} \ln \rho_{AB} &= \ln \rho_A + \ln \rho_B = a\,(E_A + E_B) + b_A + b_B \\ &= aE_{AB} + b_{AB} \end{aligned} \tag{7.7}$$

whence the additivity of $\ln \rho$ is apparent.

The probability of an infinitely large energy must be infinitely small, therefore $a < 0$.

Let us introduce the notation

$$a \equiv -1/\theta \tag{7.8}$$

If the subsystems are in equilibrium, the meaning of the quantity θ is the same as in the preceding sections (product of the temperature multiplied by the Boltzmann constant). Indeed, in an ideal gas a single molecule can be treated as a separate subsystem, and the Gibbs distribution

$$e^{-E/\theta} = \exp\left(-\sum_i \varepsilon_i/\theta\right)$$

becomes the Boltzmann distribution

$$e^{-\varepsilon/\theta}$$

corresponding to the equilibrium conditions in the gas.

Introducing the notation

$$b \equiv F/\theta \tag{7.9}$$

we find that the required distribution function finally takes the form

$$\rho\,(E) = e^{(F-E)/\theta} \tag{7.10}$$

Since $\sum t(E) = t$, the following condition is imposed on the distribution function $\rho(E)$:

$$\sum_E P\,(E) = \lim \sum_E \frac{t\,(E)}{t} = 1 = \sum_E \rho\,(E)\,g\,(E) \tag{7.11}$$

This simply means that the probability of a subsystem occurring in any of the possible states compatible with the conservation laws is unity. With the help of the normalization conditions (7.11) we can express the quantity F as a function of θ (and, as will be shown in the next section, of some of the parameters involved in the energy E). For this it is sufficient to substitute the Gibbs distribution (7.10) into (7.11) and perform the summation over all possible states. Then the factor $e^{F/\theta}$, as a constant quantity, is taken outside the summation sign, and we obtain the following equation for determining F:

$$e^{-F/\theta} = \sum_E e^{-E/\theta} g(E) \tag{7.12}$$

The expression on the right-hand side is analogous to the one in Eq. (3.3) but applies to the general case. It is also called a partition function.

The Mean Energy of a Subsystem. In the preceding sections the mean energy of a system was mostly denoted E (without the averaging symbol). The Gibbs distribution involves the energy eigenvalue of a certain state, not its mean value. Now, however, we shall use the Gibbs distribution to calculate the mean energy of a subsystem, which we must denote $\bar{E}$ to distinguish it from E. Later on the averaging symbol can again be omitted.

The general definition of the mean consists in the following. Let a quantity f assume a value f_A in any state A. Then, if the probability of the state is equal to P_A, the mean value can be denoted

$$\bar{f} = \sum_A f_A P_A \tag{7.13}$$

For example,

$$\bar{E} = \sum_E E\rho(E)\, g(E) \tag{7.14}$$

because

$$P(E) = \rho(E)\, g(E)$$

Substituting the Gibbs distribution into (7.14), we obtain the expression for the mean energy:

$$\bar{E} = \sum_E E g(E)\, e^{(F-E)/\theta} \tag{7.15}$$

Energy Fluctuations. Mean quantities to some extent characterize the state of a system in general. All statistics, not only physical statistics, make use of mean quantities for this purpose: a constant mean quantity makes it possible to estimate the order of magnitude of a variable.

However, if a variable exhibits wide scattering, the mean value does not describe it adequately.

Therefore, in addition to the mean value of the energy of a subsystem it is interesting to know its mean scattering. These two mean quantities define the state of a subsystem much better than $\bar{E}$ alone. But if we average the quantity

$$\Delta E \equiv E - \bar{E} \tag{7.16}$$

we obtain identical zero. Indeed

$$\overline{\Delta E} = \bar{E} - \bar{E} \equiv 0 \tag{7.17}$$

It is therefore expedient to average the quantity $(\Delta E)^2 = (E - \bar{E})^2$. Since $(\Delta E)^2$ is essentially a positive quantity, the deviation of E to either side of the mean makes a contribution. The required mean quantity can be written down somewhat differently:[9]

$$\begin{aligned}\overline{(\Delta E)^2} &= \overline{(E-\bar{E})^2} = \overline{E^2} - \overline{2E\bar{E}} + (\bar{E})^2 \\ &= \overline{E^2} - 2\bar{E}\bar{E} + (\bar{E})^2 = \overline{E^2} - (\bar{E})^2 \end{aligned} \tag{7.18}$$

Here we have taken advantage of the fact that the mean of a constant quantity $(\bar{E})^2$ is equal to that quantity, and also of the fact that the constant factor can be taken outside the averaging symbol:

$$\overline{\bar{E}E} = (\bar{E}\bar{E}) = (\bar{E})^2$$

The quantity $\sqrt{\overline{(\Delta E)^2}}$ is called the *absolute fluctuation* of energy. It characterizes the average deviation of the energy from its mean value. The ratio $\sqrt{\overline{(\Delta E)^2}}/|\bar{E}|$ is called the *relative fluctuation* of energy. It is a measure of the relative part of the deviation of the energy from its mean value.

The definitions of absolute and relative fluctuations retain their meaning for other quantities (as well as energy) that describe a subsystem.

Let us now apply the Gibbs distribution to the calculation of energy fluctuation in a subsystem. For this we differentiate with respect to θ the two identities from which F and $\bar{E}$ are determined:

$$\sum_E e^{(F-E)/\theta} g(E) = 1, \qquad \bar{E} = \sum_E E e^{(F-E)/\theta} g(E)$$

The first equation is the normalization condition (7.11), the second is the definition of mean energy (7.15).

The quantities E and $g(E)$ are purely mechanical and do not depend upon the statistical characteristic θ of the system. Therefore,

---

[9] Note the difference between $\overline{E^2}$ and $(\bar{E})^2$.

only F, $\bar{E}$, and, of course, θ itself, need to be differentiated with respect to θ. Thus,

$$\sum_E \left(\frac{1}{\theta}\frac{\partial F}{\partial\theta}-\frac{F-E}{\theta^2}\right) e^{(F-E)/\theta} g(E)=0 \tag{7.19}$$

$$\sum_E \left(\frac{1}{\theta}\frac{\partial F}{\partial\theta}-\frac{F-E}{\theta^2}\right) Ee^{(F-E)/\theta} g(E)=\frac{\partial\bar{E}}{\partial\theta} \tag{7.20}$$

From (7.19) we have:

$$\begin{aligned}\frac{1}{\theta}\frac{\partial F}{\partial\theta}&=\frac{F}{\theta^2}\sum_E e^{(F-E)/\theta} g(E)-\frac{1}{\theta^2}\sum_E Ee^{(F-E)/\theta} g(E)\\ &=\frac{F-\bar{E}}{\theta^2}\end{aligned} \tag{7.21}$$

Substituting this into (7.20), we find that

$$\begin{aligned}\frac{\partial\bar{E}}{\partial\theta}&=\sum_E\left(-\frac{F-E}{\theta^2}+\frac{F}{\theta^2}-\frac{\bar{E}}{\theta^2}\right) Ee^{(F-E)/\theta} g(E)\\ &=\frac{1}{\theta^2}\sum_E (E^2-E\bar{E})\, e^{(F-E)/\theta} g(E)\end{aligned} \tag{7.22}$$

The quantity $\bar{E}$ can, as a constant, be taken outside the averaging sign, and we obtain

$$\theta^2\frac{\partial\bar{E}}{\partial\theta}=\overline{E^2}-(\bar{E})^2=\overline{(\Delta E)^2} \tag{7.23a}$$

whence the relative fluctuation is

$$\frac{\sqrt{\overline{(\Delta E)^2}}}{|\bar{E}|}=\frac{1}{|\bar{E}|}\,\theta\sqrt{\frac{\partial\bar{E}}{\partial\theta}} \tag{7.23b}$$

Since the mean energy is, as an additive quantity, proportional to the number of particles in the subsystem, this quantity is inversely proportional to the square root of the number of particles.

Let us illustrate this for the case of an ideal gas.

From (2.17) $\bar{E} = 3N\theta/2$, hence the relative fluctuation is $[2/(3N)]^{1/2}$. For example, for one cubic centimetre of gas at normal conditions $N = 2.7 \times 10^{19}$, so that the relative energy fluctuation is a few parts in 10^{10}.

Most of the time the energy of 1 cm^3 of gas in an external medium differs from its mean value by just such a small fraction. Nevertheless, in subsequent statistical development it is more convenient to treat the energy of a subsystem as a slightly fluctuating rather

than a strictly constant quantity as in the case of a perfectly isolated system.

For an individual gas molecule the relative fluctuation is not, of course, a small quantity. Thus, from (2.14) and (2.15) the fluctuation of velocity is

$$\frac{\sqrt{\overline{v^2}-(\bar{v})^2}}{\bar{v}}=\left(\frac{3\theta}{m}-\frac{8\theta}{\pi m}\right)^{1/2}\left(\frac{8\theta}{\pi m}\right)^{-1/2}=\left(\frac{1.42}{8}\right)^{1/2}=0.42$$

Thus, the probability of a given value of energy in a subsystem has a very sharp maximum close to $E=\bar{E}$. The maximum is the sharper the larger the subsystem.

Entropy. Proceeding from the probability density for a subsystem, we can construct a similar function for a closed system. Using the fact that probabilities are multiplicative, that is, they can be multiplied, we can write

$$P=\prod_i P_i=\prod_i \rho_i g_i=\prod_i \rho_i \prod_i g_i \tag{7.24}$$

We now make use of the explicit form of the Gibbs distribution (7.10). Then for the product taken over all the subsystems we obtain

$$\prod_i \rho_i=\prod_i e^{(F_i-E_i)/\theta}=\exp\left[\frac{1}{\theta}\left(\sum_i F_i-\sum_i E_i\right)\right]$$
$$=\exp\left[\frac{1}{\theta}\left(\sum_i F_i-E\right)\right]$$

But if the system is large $\sum E_i=E=\text{const}$, and consequently $\prod \rho_i=\text{const}$. Thus, the probability of a certain state is proportional to the statistical weight of that state:

$$P\sim\prod_i g_i=G(E) \tag{7.25}$$

All states of a system with the same energy are equiprobable: the probability of the states is proportional to the number $G(E)$ (see Sec. 1).

It is self-evident that perfectly closed systems do not exist in nature. When we speak of a closed system, we imply one in which its subsystems come into equilibrium faster than the large system as a whole attains equilibrium with the surroundings. The time it takes for the subsystems to come into equilibrium is too small for the additive integrals of the larger system to change perceptibly. We can therefore distinguish between statistical equilibrium in the whole system and equilibrium in its subsystems.

Obviously, statistical equilibrium in a large system is maintained longer than equilibrium established only within its subsystems. That is why the probability of a fuller equilibrium is, by definition, greater than the probability of the less full equilibrium. From (7.25) the measure of probability for a large system is the statistical weight of its state. Therefore, the closer a closed system approaches statistical equilibrium, the greater the statistical weight of its state. Hence, $G(E)$ can serve as a measure of the closeness of a large system to equilibrium. Similarly, we can regard the quantity g_i of each ith subsystem as a measure of its closeness to (internal) equilibrium for the time intervals in which the subsystem can be treated as quasi-closed.

For any not too small time interval we can indicate systems that remain virtually closed during that interval. For them the quantity G is a measure of the equilibrium of their states: the greater G, the closer the subsystems of a given "closed" system are to mutual equilibrium.

Since $\ln G$ possesses the property of additivity, it is more convenient to use it rather than the statistical weight G itself as a measure of a system's closeness to statistical equilibrium. The quantity $\ln G$ is called the *entropy* of the system and is denoted S:

$$S = \ln G \tag{7.26}$$

It was shown in the preceding section that the state of a Fermi gas at absolute zero is defined solely by the fact that $G = 1$. Hence, at absolute zero ($\theta = 0$) the entropy of a Fermi gas is zero:

$$S = \ln 1 = 0$$

A Bose gas at absolute zero is completely in a zero energy state (see Sec. 5). Hence, its state at absolute zero is determined by the fact that $S = 0$.

Entropy in a Subsystem. Since

$$S = \ln G = \ln \prod_i g_i = \sum_i \ln g_i = \sum_i S_i \tag{7.27}$$

entropy is by definition an additive quantity.

This equation implies that it is natural to call the quantity $\ln g_i = S_i$ the *entropy of a subsystem*. To calculate it we shall make use of the Gibbs distribution for a subsystem. As was shown in this section, the energy of a quasi-closed subsystem is very close to its mean value $\bar{E}_i$, but not exactly equal to it.

Therefore, the formula for the entropy of a subsystem can be successfully applied to a "closed" system whose energy is exactly constant. The error here is determined by the relative fluctuation of the quantities in the subsystem and hence is negligibly small.

The entropy of a quasi-closed system, which is equal to $\ln g_i(E_i)$, should be represented as $\ln g_i(\bar{E}_i)$, where $\bar{E}_i$ is the mean value of the energy in the given subsystem in the case of a "frozen" interaction with the other subsystems. In other words, in determining $\bar{E}_i$ it is assumed that in the given time interval the subsystems do not come into more or less complete equilibrium among themselves.

Taking advantage of the fact that the energy fluctuations are small, we can replace condition (7.11) with the following simple relationship:

$$\sum_{E_i} \rho(E_i)\, g_i(E_i) \approx \rho(\bar{E}_i)\, g_i(\bar{E}_i) = 1 \tag{7.28}$$

Substituting $g_i(\bar{E}_i)$ into the definition of the entropy of a subsystem, we find that

$$S_i = \ln \frac{1}{\rho(\bar{E}_i)} \cong \ln \frac{1}{\overline{\rho_i(E_i)}} \tag{7.29}$$

But since a logarithm is a slowly varying function, we can replace the logarithm of the mean value by the mean value of the logarithm:

$$S_i = \overline{\ln \frac{1}{\rho_i}} \tag{7.30}$$

The resultant error is the smaller the larger the subsystem since the relative fluctuations tend to zero as the subsystem grows larger.

Substituting ρ_i from the Gibbs distribution (7.10) into (7.30), we obtain the following expression for entropy (omitting the subscript i):

$$S = \overline{\ln \frac{1}{\rho}} = -\overline{\ln e^{(F-E)/\theta}} = \frac{\bar{E}-F}{\theta} \tag{7.31}$$

Comparing this with (7.21), we obtain

$$S = -\frac{\partial F}{\partial \theta} \tag{7.32}$$

Substituting $\bar{E} - \theta S$ for F, we arrive at the relationship

$$S = -\frac{\partial(\bar{E}-\theta S)}{\partial \theta} = -\frac{\partial \bar{E}}{\partial \theta} + S + \theta \frac{\partial S}{\partial \theta}$$

or

$$\theta \frac{\partial S}{\partial \theta} = \frac{\partial \bar{E}}{\partial \theta}$$

Here the differentiation occurs with respect to θ under constant external conditions on which $\bar{E}$ and F may depend. The obtained relationship can also be written as

$$\theta = \frac{\partial \bar{E}}{\partial S} \tag{7.33}$$

EXERCISES

1. Using Eq. (7.26), find the expression for the entropy of an ideal monatomic gas in terms of its energy and volume. The number of atoms is N.

Solution. First compute the phase volume of all states with energies below the given energy. Dividing by $(2\pi h)^{3N}$, we obtain

$$\Gamma = \int \frac{dp_{x1}\, dp_{y1}\, dp_{z1}\, dp_{x2} \,\ldots\, dp_{zN}\, dx_1\, dy_1\, dz_1 \,\ldots\, dz_N}{(2\pi h)^{3N}}$$

$$= V^N \int \frac{d\tau_1 \,\ldots\, d\tau_N}{(2\pi h)^{3N}}$$

where $d\tau_1 = dp_{x1}\, dp_{x2}\, dp_{x3}$, etc. The linear momenta of the atoms in all states with energies less than E satisfy the inequality

$$p_{x1}^2 + p_{y1}^2 + p_{z1}^2 + p_{x2}^2 + \ldots\, p_{zN}^2 \leqslant 2mE$$

The momenta are numbered, which corresponds to nonquantum statistics. The domain of $3N$-dimensional space over which the integration is carried out is analogous to a sphere in three-dimensional space. The coordinates of points within the sphere satisfy the inequality

$$x^2 + y^2 + z^2 \leqslant R^2$$

where R is the radius of the sphere. The radius of a $3N$-dimensional sphere is equal to $(2mE)^{1/2}$. Obviously, volume in a $3N$-dimensional space is proportional to $(2mE)^{3N/2}$, just as in three-dimensional space volume is proportional to R^3. (The coefficient dependent on N is not required in this exercise.) Then the number of states between E and $E + dE$ is proportional to the quantity

$$\frac{d\Gamma}{dE} \sim (2mE)^{(3N-2)/2}\, V^N$$

Since entropy is equal to the logarithm of statistical weight, it includes the component $(3N/2 - 1) \ln E$. Neglecting unity in comparison with $3N/2$, we arrive at the equation expressing the dependence of entropy on the energy and volume of the gas:

$$S = \frac{3N}{2} \ln E + N \ln V + \text{constant}$$

2. Show that the quantum mechanical density matrix [Sec. 27] corresponds to the Gibbs distribution

$$\rho = \exp\left(\frac{F - \hat{\mathcal{H}}}{\theta}\right)$$

where $\hat{\mathcal{H}}$ is the Hamiltonian operator.

Solution. For better symmetry in writing the equations we shall assume the operator $\hat{\mathcal{H}}$ in its matrix form $\mathcal{H}_{x'x}$. Write the expression for the quantum mechanical mean of a certain quantity λ whose coordinate representa-

tion has the matrix form $\lambda_{xx'}$ [Eq. (25.19)]:

$$\langle\lambda\rangle_E = \int \psi^*(E,\ x)\,\lambda_{xx'}\psi(E,\ x')\,dx\,dx'$$

Here x is the set of coordinates of the system. The statistical mean $\bar{\lambda}$ is linked with $\langle\lambda\rangle_E$ by a relationship analogous to (7.14), so that

$$\bar{\lambda} = \sum_E \langle\lambda\rangle_E\, e^{(F-E)/\theta}$$

(degeneracy is assumed to be completely removed).

We make use of the fact that a function of E multiplied by $\psi(x)$ is equal to the result of an operator applied to $\psi(x)$, where the operator is a function of $\hat{\mathcal{H}}$ of the same form as the original function of E. Thus

$$e^{-E/\theta}\psi(E,\ x') = e^{-\hat{\mathcal{H}}/\theta}\psi(E,\ x') = \int e^{-\mathcal{H}_{x'x''}/\theta}\psi(E,\ x'')\,dx''$$

Interchanging the sum with the integral, we obtain

$$\bar{\lambda} = \int \sum_E \psi^*(E,\ x)\,\lambda_{xx'}e^{(F-\mathcal{H}_{x'x''})/\theta}\psi(E,\ x'')\,dx\,dx'\,dx''$$

But

$$\sum_E \psi^*(E,\ x)\,\psi(E,\ x'') = \delta(x-x'')$$

This is obtained by analogy with [Eq. (26.28)] if we interchange x and the eigenvalue E in the wave function according to [26.31]:

$$\psi(E,\ x) = \psi^*(x,\ E)$$

Finally, we arrive at the expression

$$\bar{\lambda} = \int dx'\,dx\,\lambda_{xx'}e^{(F-\mathcal{H}_{x'x})/\theta}$$

$$= \int dx\,(\hat{\lambda}e^{(F-\hat{\mathcal{H}})/\theta})_{xx} = \mathrm{Tr}\,(\hat{\lambda}e^{(F-\hat{\mathcal{H}})/\theta})$$

A comparison with [27.36] shows that $\rho = e^{(F-\hat{\mathcal{H}})/\theta}$ is the density matrix of a system in statistical equilibrium.

8

THERMODYNAMIC QUANTITIES

Statistics and Thermodynamics. The findings of the previous section may appear somewhat abstract if they are not related to real, measurable properties of macroscopic bodies. These are the properties that describe the behaviour of bodies in compression, heating and other

processes in terms of corresponding constants (compressibility, thermal expansion and other macroscopic characteristics of bodies). Statistics tells how to establish the relationships between these constants in general form and how to calculate their mean values determined from the Gibbs distribution or in terms of parameters entering that distribution.

A quantity such as θ (called the *distribution modulus*) is in essence defined by the way it appears in the Gibbs distribution. In calculating mean values, θ appears under the summation or integral sign as a parameter and is, therefore, one of the quantities defining the macroscopic state of a system.

The properties of mean macroscopic quantities defining the state of a body form the subject of thermodynamics. These properties are expressed in the form of a series of relationships, differential and integral, which will be developed and interpreted in this section.

Historically, thermodynamics appeared before statistics. It was conventionally based, as is known, on two postulates or laws. The validity of the laws of thermodynamics has been confirmed by a vast number of experimental facts. That is why thermodynamics can be studied irrespective of statistical physics, especially with regard to its technical applications. It should be remembered, however, that today the laws are no longer mere postulates since they are based on statistical methods.

Statistical physics is not only a justification of thermodynamics. In the first place, statistics provides a method for computing thermodynamic quantities from the microscopic structure of bodies. Furthermore, statistics makes it possible to precompute the degree to which true quantities deviate from their mean values. Such deviations, as was shown before, are measured by fluctuations of the type (7.22). In certain conditions fluctuations manifest themselves in ways that make it possible to record them experimentally (see Sec. 9).

When thermodynamics was still in the making, the atomic structure of matter had not yet been finally proved, which made postulates highly essential. But today they should not be regarded as unprovable. Any property of matter can, in principle, be deduced from elementary laws.

It should not be imagined, however, that with the appearance of statistics thermodynamics has lost its significance as a department of physics. Thermodynamics teaches how real, experimentally observable macroscopic quantities determining the thermal, mechanical, chemical and other properties of macroscopic bodies are interrelated. In those cases when the calculation of some quantity by statistical methods is practically impossible due to lack of knowledge of the elementary laws of the forces of interaction or to great mathematical complexity, thermodynamics indicates how the quantity can be determined, directly or indirectly, by measurement.

Giving preference to the systematic presentation over the historical, we shall base thermodynamics wholly on statistics. With this approach we need not invoke the laws as postulates.

Quantity of Heat. A quasi-closed macroscopic system attains a state of statistical equilibrium within itself much faster than with the surrounding medium. It spends most of the time in this state, the true values of quantities being almost constant and close to their mean values.

If two or more subsystems in internal equilibrium are brought into contact, equilibrium is established between them. The measure of a system's equilibrium is its entropy. We shall now examine how interactions between systems that result in equilibrium affect the macroscopic quantities that characterize their states.

Let two bodies be brought into contact in such a way that the external conditions and the number of particles in each of them remain unchanged. Then, from the differential equation (7.33), the mean energy increment of each subsystem is proportional to its entropy increment:

$$dE = \theta\, dS \tag{8.1}$$

Here the partial differentials are replaced by total differentials, keeping in mind the conditions of obtaining them. The average symbol over E has been omitted since thermodynamics always deals with mean quantities and only they can appear in equations.

The total energy increment for the two bodies isolated from external action is zero:

$$dE_1 + dE_2 = 0 \tag{8.2}$$

The total entropy increment is positive or zero, because as a result of the interactions the bodies come into mutual statistical equilibrium. Also, this equilibrium is more complete than the equilibria inside each of them. Hence,

$$dS_1 + dS_2 \gg 0 \tag{8.3}$$

Using (8.1) and (8.2), we obtain

$$dE_1\left(\frac{1}{\theta_1} - \frac{1}{\theta_2}\right) \geqslant 0 \tag{8.4}$$

If $\theta_1 > \theta_2$, then $dE_1 < 0$, that is, the first system transmits energy to the second. The transmission of energy takes place entirely due to contact interaction, that is, to the microscopic forces between molecules at the points of contact. Energy thus transferred is termed *heat*; it follows that heat should not be called a "form of energy". Rather, it would be correct to speak of heat as a mode of energy transfer from one body to another.

7—0493

In formula (8.4), θ_1 and θ_2 are parameters involved in the Gibbs distributions of each subsystem separately. As long as these parameters differ the systems cannot be in equilibrium between themselves. Approximation to equilibrium occurs as heat is transferred, the transfer always occurring in the direction of the subsystem with the smaller value of θ. Only when θ_1 and θ_2 are the same does the transfer of macroscopic quantities of heat cease, and the energy of each subsystem experiences only small fluctuations around the equilibrium value. If one of the systems is an ideal Boltzmann gas, then, as was shown before, θ is proportional to the absolute temperature, since the Gibbs distribution for a gas as a whole leads to the Boltzmann distribution for the individual molecules with the same parameter θ. The absolute temperature of a gas can be determined by independent, nonthermal measurements from the ideal gas law $pV = RT$. It is natural to regard the quantity θ for any system that is not an ideal gas as simply temperature. If a system is in equilibrium with an ideal gas its value of θ is the same as for the gas and, hence, proportional to its absolute temperature. Thus, the quantity θ in the Gibbs distribution of a quasi-closed subsystem has, in fact, the meaning of its temperature, measured in absolute units (ergs), if the ideal gas is taken as a *thermometric* substance. A definition of temperature that does not depend on the choice of thermometric substance will be given later on in this section.

The Gibbs distribution occurs for any assembly of quasi-independent subsystems, including those that have not yet arrived at a state of mutual statistical equilibrium. Although in this case the quantity θ is by definition the same for all subsystems, which follows from the multiplicativity of the distribution function $\rho(E)$ (see (7.4)-(7.8)), it cannot be regarded as equal to the temperature θ of the large system. A system that is not in equilibrium does not, generally speaking, have an exactly defined temperature. If the subsystems are only in internal equilibrium, each is characterized by its own Gibbs distribution, neither of which can be taken as a factor in the Gibbs distribution of the larger system since the parameters of the system and subsystem are different. They coincide only when equilibrium exists, in which case they are a measure of the temperature of the system.

The example of temperature shows how statistically defined quantities are identified with directly measurable thermodynamic quantities. A statistical quantity can be regarded as defined if and only if there is given a unique group of operations (of measurement and calculation) relating this quantity to real macroscopic quantities or to experimentally measurable microscopic parameters of a system.

Work. The Hamiltonian function or the Hamiltonian operator of a system usually depends not only on generalized coordinates and

momenta that vary according to the laws of dynamics but also on certain arbitrarily chosen parameters, for example, the intensity of an external electromagnetic field. The energy spectrum, and hence the mean energy E of the system, depends upon the parameters appearing in the Hamiltonian.

These arbitrarily variable parameters are called the *external parameters* of a system. We denote them, for the most general case, by the letter λ, where λ may mean any quantity of this type. As λ varies the mean energy also varies. Obviously, it can vary only through the action of some external energy source. Since λ is a mechanical and not a statistical quantity (it is involved in the Hamiltonian!), the variation in λ is due to some external mechanical work done on the system, for example, a falling weight or a rotating motor.

The mechanical work done in the changing of λ can be represented as

$$dA = -\Lambda \, d\lambda \tag{8.5}$$

where it is natural to call the quantity Λ the generalized force (since work is equal to the product of the "force" Λ and "path" $d\lambda$). The minus sign has been introduced into the equation to equate the work dA to the energy increment of the body on which the work is done. Indeed, the energy increment of the body is by definition the quantity

$$dE = \overline{\frac{\partial E}{\partial \lambda}} \, d\lambda \tag{8.6}$$

where on the left the averaging symbol is implied and has been deleted solely to preserve a unified notation in this section. Comparing (8.5) and (8.6), we see that the mean quantity $\overline{\partial E/\partial \lambda}$ is equal to the generalized force taken with opposite sign:

$$\Lambda = -\overline{\frac{\partial E}{\partial \lambda}} \tag{8.7}$$

Thus, we have obtained a generalized relationship between force and energy of the type [2.1], as could have been expected.

In practical applications the most frequently employed external parameter of a system is the volume it occupies. In purely mechanical terms this can be visualized by assuming that the potential energy of any particle belonging to the system is equal to infinity beyond the boundaries of the volume, that is, that infinite work is required to remove even a single particle from the volume. This is how the volume appears in the Hamiltonian of the system (see [Sec. 28]).

Imagine a system occupying the volume of a cylinder with a movable piston. Denoting the pressure p and the area of the piston f, we find that the force acting on the piston is pf. The work done on the piston in its displacement through a distance dx is $pf\,dx$. The

work done on the system is, accordingly, $dA = -pf\,dx$. But the product $f\,dx$ is equal to the volume increment dV of the system. Hence, the change in energy of the system is

$$dE = dA = -p\,dV \tag{8.8}$$

In compression ($dV < 0$) the work is positive.

It can be seen from Eq. (8.8) that the pressure is the generalized force Λ related to the volume increment dV.

Thus, the energy of a system may vary when external parameters vary. In thermodynamics this mode of energy variation is called *work*, thereby generalizing the mechanical concept of work.

The First Law of Thermodynamics. As we have shown, energy can be transmitted to a system by purely contact action, without any change in macroscopic parameters or particle exchange. This type of energy transmission was called *heat transfer*. Thus, the total change in the energy of a system consists in the work done on the system and the quantity of heat transferred to it:

$$dE = dA + dQ \tag{8.9}$$

The quantity on the left-hand side is, of course, the mean energy of the system. Equation (8.9), which expresses the law of conservation of energy, can also be considered as an identity defining the quantity of heat: $dQ = dE - dA$. Proceeding from the statistical interpretation of thermodynamics, we can be sure that the energy conservation law is applicable to thermal processes. Any energy imparted to a system without altering its external parameters must be transmitted through contact; it is this energy that has been called the *quantity of heat*.

But thermodynamics appeared before statistics. Interpreted thermodynamically, Eq. (8.9) means that quantity of heat can be measured in units of mechanical work, and work can be measured in units of quantity of heat. In other words, Eq. (8.9) extends the energy conservation law to thermal processes. That is why the establishment of the mechanical equivalent of heat by Julius R. von Mayer, James P. Joule and Hermann von Helmholtz represented a major breakthrough in the advance of physical knowledge. The earlier view of heat as latent motion of molecules, although close to the modern statistical interpretation of thermal phenomena, contained no quantitative relationships. Therefore, the theory of heat and, especially, heat engines, could develop only after the correspondence between thermal and mechanical quantities was proved experimentally. Only after the fundamental propositions of thermodynamics were formulated did statistics begin to develop as a physical, quantitative theory.

Equation (8.9) can be reduced to another form. For this we must note that the energy of a body is a single-valued function of its

state. Imagine a certain alternating process in which heat is supplied to a body, and the body delivers work, as is the case in heat engines. Integrate Eq. (8.9) over one operating cycle:

$$\int dE = \int dQ + \int dA \tag{8.10}$$

The energy has the same value at the beginning and end of the cycle, which is, in fact the periodicity condition. Therefore, the total energy change $\int dE$ over one cycle is zero. Hence,

$$\int dQ = -\int dA \tag{8.11}$$

The work done by the engine over one cycle is equal to the quantity of heat delivered to it in that cycle. It is impossible to build an engine capable of working without an external supply of heat (or energy in general). This statement is called the "First Law of Thermodynamics". An imaginary engine performing work without an external energy source is called a *perpetual motion engine of the first kind.* The inevitable failure of all attempts to build such an engine ultimately led to the negative postulate on which thermodynamics was based. Of course, if thermodynamics is based on a statistical interpretation, the first law follows from the mechanical law of conservation of energy.

Neither work nor quantity of heat, taken separately, can characterize the *state* of the body to which they are transferred. From Eq. (8.11), a body can perform any number of operating cycles, reverting every time to its initial state. In the process it receives any amount of heat and does any amount of work, reverting after each cycle to its initial state. It is therefore wrong to speak of a body's "latent heat". All it possesses is latent energy, which varies with the transmission of heat and performance of work. It is incorrect to call heat and work "forms of energy": they are but different modes of transmitting energy, one microscopic, the other macroscopic. This is seen mathematically in the fact that dA and dQ are not total differentials of any quantities. For example, $dA = -p\,dV$. Pressure depends not only on volume but on temperature as well. Thus, for an ideal gas $p = N_A\theta/V$, hence $dA = -(N_A\theta/V)\,dV$. This equation cannot be integrated until we know the temperature as a function of the volume V in the given process. Thus, heat and work characterize a process performed by a body. They do not characterize the state of the body.

In certain cases the heat transferred in a process can be expressed very simply. If, for example, the volume of a body does not change (an *isochoric* process), then $dV = 0$. In general, $dA = 0$ if the external parameters λ are constant. In that case the quantity of heat

equals the change in energy of the body:

$$dQ = dE, \qquad Q = \Delta E \tag{8.12}$$

If the pressure does not change (an *isobaric* process), then $dA = -p\,dV = -d\,(pV)$ and

$$dQ = dE + d\,(pV) = d\,(E + pV)$$

The quantity

$$E + pV \equiv H \tag{8.13}$$

is, like energy, uniquely defined by the body's state. It is called the *heat content* or *enthalpy* of the body and is denoted by the letter H. Thus, in an isobaric process the quantity of heat is equal to the change in the body's enthalpy:

$$dQ = dH, \qquad Q = \Delta H \tag{8.14}$$

Reversible Processes. To each value of the external parameters describing a subsystem of a closed system there corresponds a definite state of statistical equilibrium. We can, for example, visualize a substance behind a piston in a cylinder that is not thermally isolated. The substance and the surroundings should in such a case be treated as one system. The external parameter defining the state of the system is, in this case, the volume V occupied by the substance.

For every value of the volume a state of statistical equilibrium is established between the substance and the surroundings, when the temperatures of the surroundings and the substance are the same and the total entropy has a maximum corresponding to a given value of the total energy and the volume V behind the piston.

Let us suppose that the external parameter λ varies so slowly that for every value of λ full equilibrium has time to set in. In other words, the state of the system depends only upon the value of λ at the given time. To this value of λ corresponds the maximum entropy, so that the system is all the time in a state of statistical equilibrium. But it follows from this that in such a process the system never approaches statistical equilibrium for the simple reason that it is never brought out of that equilibrium. And since entropy is the measure of equilibrium, it does not change with such a slow variation of λ. We are speaking, of course, of the entropy of the whole system, not of its subsystems.

We can demonstrate this by the following simple reasoning. Let the rate of change of λ be $\dot{\lambda}$. By definition this is a small quantity. The rate of change of entropy is $-\dot{S}$. At $\dot{\lambda} = 0$, $\dot{S} = 0$. Let us express $\dot{S}$ in terms of $\dot{\lambda}$. It is immediately apparent that the expansion can only have the form $\dot{S} = a(\dot{\lambda})^2$. Indeed, whatever changes leading

to equilibrium the system undergoes, its entropy can only increase, so that $\dot{S} > 0$. The derivative $\dot{\lambda}$, however, varies arbitrarily and can be of both signs. Hence, the expansion begins with the term quadratic in $\dot{\lambda}$, and the change in entropy is of the second order of smallness.

The constancy of entropy for slow variations of λ can be explained in the following way. Entropy is the logarithm of the number of equiprobable states of a system in a certain range of energy values close to E. If λ varies very slowly, the entire large system must at each instant of time be regarded as conservative, so that all its individual states are equally probable. A rapid change could induce transitions in some definite direction and thereby disturb the equiprobability of states that follows from the principle of detailed balance, that is, the equiprobability of direct and reverse transitions. Since λ is a parameter involved in the Hamiltonian, the total number of states is conserved for slow variations of λ. Only the degeneracy of states may depend on it but not their number. The most probable range of states having equal probability of occurrence is, in principle, determined purely by combinatorial analysis and therefore does not depend upon the particular value of λ for which the states are taken. Consequently, the number of states in the most probable region and the logarithm of that number, that is, the entropy, are conserved.

We have thus shown that to each value of λ, when the variation is slow, there corresponds a definite state of the system, regardless of the way in which the value of λ varied prior to that, provided the variation was slow enough. Let λ change first from λ_1 to λ_2, and then from λ_2 to λ_1. Then in the latter process the system will pass through the very same states it occupied when λ was varying in the former. Such processes are called *reversible.*

We can imagine the following two limiting cases.

(1) A subsystem and the surrounding medium are continuously in statistical equilibrium, so that their temperatures are the same. If the surroundings are sufficiently large, their temperature does not change at all, and consequently, the temperature of the subsystem does not change either in the process. Such a reversible process is said to be *isothermal.* In an isothermal process the entropy of the total system is conserved, while the entropy of the subsystem and the entropy of the surroundings vary by amounts equal in absolute value but of opposite sign.

(2) The parameter λ varies so fast that an approach to statistical equilibrium between the surroundings and the subsystem does not have time to occur, but at the same time the variation is so slow that the equilibrium within the subsystem and within the surroundings is not disturbed. Such a process would occur if the system were

separated from the surroundings by an ideal thermal insulation. As heat transfer is usually a slow process, we can readily imagine such rapid variations of λ that there is not enough time for heat to be transferred. In this process, the entropy of the system and the entropy of the surroundings is conserved separately, because the variation of λ is slow with respect to the equilibrium mechanism. Such a process is called *isentropic* or *adiabatic*.

Further on certain irreversible processes will be examined.

The Second Law of Thermodynamics. Let us find an expression for the quantity of heat received by a system in a reversible process. As usual, we shall regard the given system as a subsystem of some larger closed system. The state of such a *quasi-equilibrium* subsystem at every given instant is fully defined by its entropy and external parameters. From (8.1) and (8.6), the energy increment for a constant number of particles is expressed in terms of the entropy increment in the following way:

$$dE = \theta\, dS + \overline{\frac{\partial E}{\partial \lambda}}\, d\lambda \tag{8.15}$$

Applying (8.5) and (8.7) to this equation, we obtain

$$dE = \theta\, dS - \Lambda\, d\lambda = \theta\, dS + dA \tag{8.16}$$

whence it follows that

$$\theta\, dS = dE - dA \tag{8.17}$$

But the right-hand side of the last equation is the quantity of heat received by the system, dQ. Hence, in a reversible process

$$dQ = \theta\, dS \tag{8.18}$$

Irreversible processes may be taking place in other subsystems of the larger system to which the subsystem under consideration belongs, but this does not affect the applicability of Eq. (8 18). It is one of the most important equations of thermodynamics and defines the entropy increment of a system in terms of quantity of heat, which is directly measured by experiment. It is most significant that the quantity of heat received by a system in some reversible, or in general any, process depends upon the development of the process, whereas the entropy increment is determined only by the initial and final states of the system. The ratio of an infinitely small quantity of heat received by a subsystem in a reversible process to the temperature is a total differential:

$$dS = \frac{dQ}{\theta} \tag{8.19}$$

If an irreversible process occurs within the subsystem, Eq. (8.19) may not hold. Indeed, let a system consist of two conjugate sub-

systems at different temperatures. In the process of temperature equalization such a system approaches statistical equilibrium and its entropy increases. But no heat reaches the system from outside, so that dQ for the system as a whole is zero, and $dS > 0$.

Here is another example of an irreversible process. Let a gas initially contained in a vessel of volume V_0 be passed through an orifice into an evacuated vessel, so that at the end of the process it occupies a greater total volume V. The phase volume Γ naturally increases, since the geometrical volume increases (see Exercise 1, Section 7). But this means that the entropy also increases. When expanding into a vacuum a gas does not perform work (since there are no opposing forces) and does not receive heat. In other words, its energy is conserved (see Eq. (8.9)) and it can therefore be regarded as a closed system approaching statistical equilibrium. Obviously, if one of two communicating vessels is evacuated, the system is not in equilibrium and its entropy increases. Note that when a gas expands isothermally in a cylinder with a piston subject to the pressure of the surroundings, the entropy of the gas also increases, but the entropy of the surroundings decreases to the same extent. Thus, the entropy increment in an irreversible expansion of gas into vacuum is positive, and the transferred quantity of heat is zero.

The two foregoing examples show that if a process takes place *within* a system, then

$$\frac{dQ}{\theta} \leqslant dS \tag{8.20}$$

If a given system irreversibly exchanges heat with other systems and no irreversible processes take place *within* it, then Eq. (8.19) is applicable.

Let us now use Eq. (8.19) to determine the work that can be performed by a heat engine. By this term we mean a device that periodically receives heat from some heat reservoir and use it to perform work. According to the first law of thermodynamics, the total work done in one operating cycle is equal to the quantity of heat received in that cycle (8.10). If the engine operates reversibly, the quantity of heat is given by Eq. (8.18). Therefore

$$\int dA = -\int dQ = -\int \theta \, dS \tag{8.21}$$

It follows from this that if the temperature of the working medium remains unchanged in the course of a cycle, the work is identically zero:

$$\int dA = -\theta \int dS = 0 \tag{8.22}$$

(in a periodic process the initial state coincides with the final state, and the entropy is a single-valued function of the state, so that

$\int dS = 0$). In irreversible processes

$$dQ < \theta\, dS, \qquad \int dQ > 0$$

so that if the temperature is constant, $\int dA \geqslant 0$. In that case a periodic process can be sustained only by external work done on the system.

It follows from Eq. (8.22) that a heat engine cannot function if it receives heat only from the surroundings since they are, by definition, at constant temperature. The statement formulated here is known as the "Second Law of Thermodynamics".

An imaginary engine designed to operate solely from heat derived from the surroundings is called a *perpetual motion engine of the second kind*. In the axiomatic presentation of thermodynamics the impossibility of building such an engine was postulated (on the basis of countless and fruitless attempts to construct it), and the subsequent proofs were indirect: first it was assumed that the statement to be proved was false which, if so, meant that a perpetual motion of the second kind could in fact be built. As entropy is a statistical concept, a perpetual motion engine of the second kind should be defined as an infinitely improbable device, and a perpetual motion engine of the first kind as a mechanically impossible system violating the energy conservarion law.

A perpetual motion engine of the second kind should not be confused with a so-called *free* engine, like a wind motor, which functions because of the sun's heating of the earth.

Efficiency. For a heat engine to work it must have an operating cycle passing through two temperatures. The higher temperature is conventionally called the *temperature of the heat source*, the lower is called the *temperature of the heat sink*. The work done in one cycle is

$$\int dA = \theta_1 \int_a^b dS + \theta_2 \int_b^a dS \tag{8.23}$$

where the limit a refers to the initial and final states, and the limit b refers to an intermediate state. But $\int_b^a dS = - \int_a^b dS$, so that

$$\int dA = (\theta_1 - \theta_2) \int_a^b dS \tag{8.24}$$

The total quantity of heat supplied by the source is $\int dQ = \theta_1 \int\limits_a^b dS$

The *efficiency* η of an engine is the term used for the ratio of the work done by it to the quantity of heat taken from the source, since the main losses are associated with obtaining this heat. From Eq. (8.24), the efficiency of a reversible engine is

$$\eta = \frac{\int dA}{\int dQ} = \frac{\theta_1 - \theta_2}{\theta_1} = 1 - \frac{\theta_2}{\theta_1} \tag{8.25}$$

The equation shows that the efficiency of a reversible engine depends upon the temperatures of the source and the sink. The temperature θ_2 is actually either the temperature of the surroundings or a somewhat higher temperature. To increase efficiency we must increase θ_1.

Equation (8.25) shows that the efficiency of a reversible engine can be used to define the *absolute thermodynamic temperature scale* independent of the thermometric substance. But, as can be seen from (8.25), this scale coincides with the scale of the gas thermometer, which explains the special importance of the latter.

The efficiency of an irreversible engine is less than that of a reversible engine operating at the same source and sink temperatures. Indeed, when (8.20) is taken into account, Eq. (8.24) is replaced by the inequality $\int dA \leqslant (\theta_1 - \theta_2)(S_a - S_b)$. That is why, given the same quantity of heat taken from the source, the work done by an irreversible engine is less than that done by a reversible engine.

The efficiency of an irreversible engine is less because part of the heat obtained from the source is wasted on overcoming friction or is dissipated into the surroundings as, for example, through the cylinder's walls of a piston engine.

It should be noted that a perfectly reversible engine would have to operate infinitely slowly, since otherwise there would be no time for statistical equilibrium to be established at every instant. Approximation to equilibrium is always irreversible.

The Differential Thermodynamic Identities for Energy and Enthalpy. Proceeding from the general equation (8.9), we can write a general equation for the differential of the mean energy of a system in the case of a constant number of particles, taking volume as the external parameter:

$$dE = \theta\, dS - p\, dV \tag{8.26}$$

In this formula dS denotes the entropy increment due to the reversible processes in the subsystem and interaction with the surroundings. Remember that in an open system the change in entropy is not

associated with approximation to equilibrium. In particular, dS can be of either sign. The state of a homogeneous system with a constant number of particles is defined by two quantities: volume and entropy. This can be seen from the number of independent parameters appearing in the Gibbs distribution: S and $V = \lambda$ can be taken instead of θ and $V = \lambda$; the energy of such a system depends on entropy and volume. Let us find the total differential of this function:

$$dE = \left(\frac{\partial E}{\partial S}\right)_V dS + \left(\frac{\partial E}{\partial V}\right)_S dV \tag{8.27}$$

where the subscripts denote the quantities that remain fixed during differentiation. Comparing (8.26) and (8.27), we have

$$\theta = \left(\frac{\partial E}{\partial S}\right)_V, \qquad p = -\left(\frac{\partial E}{\partial V}\right)_S \tag{8.28}$$

Differentiating θ with respect to V, and p with respect to S, we obtain an equation between the mixed partial derivatives:

$$\left(\frac{\partial \theta}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V = \frac{\partial^2 E}{\partial V\, \partial S} \tag{8.29}$$

Enthalpy is connected with energy through the relationship (8.13):

$$E = H - pV$$

whence we obtain the expression for the total differential of enthalpy:

$$dH = \theta\, dS + V\, dp \tag{8.30}$$

It is assumed here that enthalpy is an explicit function of entropy and pressure, just as energy is expressed in terms of entropy and volume in the identity (8.26). The identity for enthalpy leads to a series of differential relationships:

$$\theta = \left(\frac{\partial H}{\partial S}\right)_p, \quad V = \left(\frac{\partial H}{\partial p}\right)_S, \quad \left(\frac{\partial \theta}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p = \frac{\partial^2 H}{\partial p\, \partial S} \tag{8.31}$$

The identities obtained make it possible to calculate various thermodynamic quantities in terms of others.

Free Energy. If an irreversible process takes place in a system, then from (8.20), $dQ \leqslant \theta\, dS$. Substituting this inequality into the equation of the first law of thermodynamics (8.9), we obtain

$$dE \leqslant \theta\, dS + dA \tag{8.32}$$

Thus, the work done on the system satisfies the inequality

$$dA \geqslant dE - \theta\, dS \tag{8.33}$$

Let the process take place at constant temperature. Then (8.33) can be written down as a relationship between total differentials:

$$dA \geqslant d\,(E - \theta S) \tag{8.34}$$

The quantity

$$E - \theta S \equiv \overline{E} - \theta S \equiv F \tag{8.35}$$

appears, by (7.31), in the Gibbs distribution (7.10); it is called the *free energy* of the system (or the *Helmholtz free energy*, to avoid confusion with another function, the *Gibbs free energy*).

It follows from the inequality (8.34) that the least work that must be done on a system at constant temperature to cause a change of state is equal to the free energy increment:

$$A_{\min} = F_2 - F_1 \tag{8.36}$$

The minimum work is required in a reversible process.

Inequality (8.34) can also have a somewhat different meaning. It defines the maximum work done by the system itself in a given change of state:

$$A_{\max} = F_1 - F_2 \tag{8.37}$$

The total entropy of the system and the surroundings is conserved in these processes, and inequality (8.32) becomes an equation.

Consider the following example. Let an ideal gas expand into vacuum. It does no work in the process, and its energy is conserved. But the energy of an ideal gas depends only upon its temperature, and not its volume. Therefore, the temperature does not change when the gas expands in vacuum. The entropy of the gas, as we have seen, increases. Hence, the minimum work required to revert the gas to its initial volume at the same temperature is equal to the change in the free energy of the gas in the expansion. Unlike the total energy of the gas, the free energy decreases when a gas expands into vacuum ($\Delta F < 0$ as $\Delta S > 0$).

It is easy to obtain the thermodynamic identity for free energy. Differentiating the relationship between the total and the free energy and substituting the identity (8.26), we obtain

$$dF = -S\,d\theta - p\,dV \tag{8.38}$$

Transformations from (8.26) to the identity (8.38) of the type (8.35) were carried out in [Sec. 10] in canonical transformations to other variables. From (8.38) we obtain the differential relationships

$$p = -\left(\frac{\partial F}{\partial V}\right)_\theta, \quad S = -\left(\frac{\partial F}{\partial \theta}\right)_V,$$

$$\left(\frac{\partial p}{\partial \theta}\right)_V = \left(\frac{\partial S}{\partial V}\right)_\theta = \frac{\partial^2 F}{\partial \theta\,\partial V} \tag{8.39}$$

These equations are especially convenient because they involve volume and temperature as the independent variables, and both can be measured directly. At the same time, the identity for energy involves entropy as an independent variable, which must itself be calculated (for example, by integrating (8.19)).

From (7.12), the free energy F is expressed in terms of a partition function:

$$F = -\theta \ln \sum e^{-E/\theta} \tag{8.40}$$

where E is the actual, not mean, energy.

The right-hand side of this equation is expressed in terms of temperature and the external parameters involved in the eigenvalues E. But θ and λ are the very variables that enter the identity (8.38). Therefore, for the determination of all thermodynamic quantities it is sufficient to calculate the partition function $\sum e^{-E/\theta}$. The actual calculation of this sum for an arbitrary system involves enormous mathematical difficulties. It has been calculated only for ideal gases and crystals, as well as for systems that closely approximate the ideal. It should be noted that even if someone should manage to evaluate the statistical sum for some specific substance, for example water, the thermodynamic laws obtained with such great difficulty would apply to water alone and not to liquids generally. The properties of ideal gases and crystals, however, follow from statistics in a very general way.

The Thermodynamic Potential. Let us now determine the minimum work required to cause a given change in a system at constant temperature and pressure. Let the temperature and pressure in the system be the same as in the surroundings. We note that in a homogeneous system with a constant number of particles, in which no phase or chemical changes occur, the state is completely defined by the pressure and temperature since the thermodynamic identities for such systems involve only two independent variables, and specification of two quantities is sufficient to determine all the others. If, however, a system consists of two phases of the same substance (for instance, liquid and vapor), the ratio between the liquid and vapor portions at the given temperature and pressure may be quite arbitrary.

As the volume of the system increases, it performs work. One can visualize, for example, a system in a cylinder with a piston whose rod is connected to some object capable of transforming only its mechanical energy, like a flywheel or load. In addition, when the system expands, work is done on the surroundings. In compression, the work dA' is done on the system. It consists of two components: the work dA obtained from the mechanical object and

the work $-p\,dV$ done by the surroundings, where p is the pressure in the surroundings, which in this process is equal to the pressure in the system. In compression, $-p\,dV$ is a positive quantity, as it should be. Since by definition the temperature of the system does not change, we can, in accordance with (8.33), write down the following relationship:

$$dA - p\,dV = dA - d\,(pV) \geqslant d\,(E - \theta S)$$

or

$$dA \geqslant d\,(E - \theta S + pV) \tag{8.41}$$

The quantity $E - \theta S + pV$ is, obviously, a function of the state of the system. This function is called the *thermodynamic potential* (or the *Gibbs free energy*) and is denoted by G:

$$G \equiv E - \theta S + pV \tag{8.42}$$

Its increment in some reversible process is equal to the minimum work that has to be done on the system at constant temperature and pressure equal to the temperature and pressure of the surrounding medium in order to change the state of the system in a given way:

$$A_{\min} = G_2 - G_1 \tag{8.43}$$

Such work is done in a reversible process. It is equal to the work, taken with the opposite sign, which the system could do on an external object under the same conditions in its transition from state *2* to state *1*. When G attains the minimum, the system is no longer capable of performing work. This, as usual, is the equilibrium condition. In the present case equilibrium should be interpreted in the thermodynamic sense, for example, with respect to a phase transition or chemical transformation. Hence, the equilibrium condition in systems capable of changes of this type is that the thermodynamic potential should be minimal.

Let us now find the thermodynamic relationships for G. From (8.42),

$$G = F + pV \tag{8.44}$$

Differentiating and substituting dF from (8.38), we obtain

$$\begin{aligned} dG &= dF + p\,dV + V\,dp = -S\,d\theta - p\,dV + p\,dV + V\,dp \\ &= -S\,d\theta + V\,dp \end{aligned} \tag{8.45}$$

Whence it follows, in the familiar fashion, that

$$S = -\left(\frac{\partial G}{\partial \theta}\right)_p, \quad V = \left(\frac{\partial G}{\partial p}\right)_\theta,$$

$$\left(\frac{\partial S}{\partial p}\right)_\theta = -\left(\frac{\partial V}{\partial \theta}\right)_p = -\frac{\partial^2 G}{\partial p\,\partial \theta} \tag{8.46}$$

The thermodynamic potential depends only upon quantities that characterize the state of a body: its temperature and pressure. At the same time G is, of course, an additive quantity: if two equal volumes of the same substance are joined at the same temperature and pressure, the total thermodynamic potential will be double that of each volume separately (see (8.42)). And since these volumes contain equal numbers of molecules, we can write

$$G = N\mu(p, \theta) \tag{8.47}$$

where μ is the thermodynamic potential related to a single molecule of the substance. The quantity μ is called the *chemical potential* of the given substance. It is shown later that it is identical with the parameter μ involved in the energy distribution of the molecules of ideal gases (see Sec. 1). It is obvious that

$$\mu = \left(\frac{\partial G}{\partial N}\right)_{p,\,\theta} \tag{8.48}$$

If a system consists of molecules of several types, for example, a solution of one substance in another, or a mixture of gases, the state is determined not only by the temperature and pressure but also by the concentrations of the substances. The concentration of the ith substance in a mixture is

$$c_i = N_i/\sum_k N_k \tag{8.49}$$

The chemical potential of the ith substance in a mixture is expressed by analogy with (8.48):

$$\mu_i = \left(\frac{\partial G}{\partial N_i}\right)_{p,\,\theta,\,N_{k\neq i}} \tag{8.50}$$

where μ_i is a function of p, θ and all the concentrations c_1, c_2, . . ., c_k, ...

Regarding the N_i as variables, we can write the total differential dG as follows:

$$dG = -S\,d\theta + V\,dp + \sum_i \mu_i\,dN_i \tag{8.51}$$

This equation generalizes (8.45) for the case of a variable number of particles.

Since the transition from E to F, G, and H does not involve the number of particles N_i, we can similarly generalize the differential relationships (8.26), (8.30), and (8.45): all we have to do is add the sum $\sum \mu_i\,dN_i$ to the right-hand side. For example, for dE we have

$$dE = \theta\,dS - p\,dV + \sum_i \mu_i\,dN_i \tag{8.52}$$

For the case of a constant volume and one type of molecules this equation reduces to the form

$$dE = \theta\, dS + \mu\, dN \tag{8.53}$$

Compare it with (1.18). The quantity S there is the entropy of the gas, as it denotes the logarithm of the number of states at a given energy. It is clear from this that μ is the chemical potential determined from (8.48).

The Joule-Thomson Effect. Let us examine one important example of an irreversible process. Let a gas be contained in a cylindrical vessel divided by a porous heat-insulating partition. The pressures on either side of the partition are different. The temperatures are different too, of course, as no heat exchange takes place through the partition. However, because of friction the gas seeping through the tiny pores dissipates its kinetic energy of directed flow, which transforms into internal energy: the kinetic energy of motion of individual molecules. The transition from orderly motion to random (thermal) motion is essentially an irreversible process called in the present case the *Joule-Thomson effect.*

We shall now show that the enthalpy H is conserved in this process. For this we must investigate the energy balance of a certain mass of gas, for example, one mole. Since the partition is heat-insulating, the change in energy of the gas is equal to the work done on it. We shall use the subscript 1 for quantities relating to the state of the gas with higher pressures, and the subscript 2 for quantities with lower pressures. Let a piston force one mole of the gas of volume V_1 at pressure p_1 through the partition; on the other side it assumes a volume V_2 at pressure p_2. For the gas of volume V_1 to penetrate the partition, work equal to p_1V_1 must be done on it. Since it emerges from the partition, the gas itself will perform work p_2V_2. Hence, the balance equation is

$$E_2 - E_1 = p_1V_1 - p_2V_2 \tag{8.55}$$

Thanks to the heat-insulating property of the partition the gas can acquire or release energy only in the form of work. Assembling the quantities with the same subscripts on one side of the equation, we obtain

$$E_1 + p_1V_1 = E_2 + p_2V_2$$

or, from (8.13),

$$H_1 = H_2 \tag{8.56}$$

Entropy in Classical and Quantum Statistics. Let us compare the definitions of entropy based on the classical and the quantum laws of motion. In the latter case entropy is defined as the logarithm

of the number of states of a system at a certain energy value. In passing to a quasi-classical approximation, the number of states of the system is equal to the phase volume $\Delta\Gamma$ it occupies divided by $(2\pi h)^n$, where n is the number of degrees of freedom. The logarithm of this ratio is entropy. Before the quantum laws of motion were discovered, entropy was defined as the logarithm of a concrete number $\Delta\Gamma$. In this definition, entropy depends on the choice of units. If, for example, the unit of mass is changed by a factor of two, then to the entropy must be added $n \ln 2$. Since the units are arbitrary, it follows that in applying the classical method of counting states entropy could be determined only to the accuracy of an arbitrary additive constant. Only the *change* in entropy in this or that process had strict meaning.

In the quantum method of counting states entropy is equal to the logarithm of a dimensionless number and does not depend on the choice of units of measurement.

The temperature of a system is equal to absolute zero when the system is in the ground state, that is, when it has the least possible energy. If such a state possesses unit weight, the entropy (or the logarithm of the weight) becomes zero. This statement is called *Nernst's heat theorem*, which is also known as the "Third Law of Thermodynamics".

Certain corollaries of Nernst's heat theorem will be examined later on.

Configurational Entropy. In some cases a considerable contribution to the entropy of a system is made by the existence of a large number of states of the system differing in the spatial configurations of its atoms or molecules. If such states have approximately the same energy, the logarithm of the number of all the states includes as a term the logarithm of the number of all spatial configurations. When the temperature is sufficiently high, so that small differences between the energies of different configurations can be neglected, their number is determined by purely combinatorial methods.

Such a contribution to the entropy may arise from a partially random distribution of atoms in a crystal that retains its structure as a whole. Let us consider a crystal of ice as an example. Like water, ice consists of individual molecules of H_2O which retain their individuality: it is always possible to indicate the water molecules in a crystal of ice to which a given atom belongs. The oxygen atoms form a regular lattice, with two, and only two, atoms of hydrogen associated with each of them. The forces that keep the water molecules in the lattice are in this case called *hydrogen bonds*. They link two atoms of oxygen each, between which there is one hydrogen atom. The oxygen atom lying closer to the hydrogen atom is connected with it by an ordinary chemical bond, the more

distant atom is connected by a hydrogen bond. It is denoted symbolically by a broken line, so that a hydrogen atom may lie between two oxygen atoms according to one of two configurations: O---H-O or O-H---O. We shall not concern ourselves with the mechanism of the hydrogen bond.

Every oxygen atom in the lattice has two "solid" and two "broken" bonds. They must be visualized spatially. Namely, each O atom

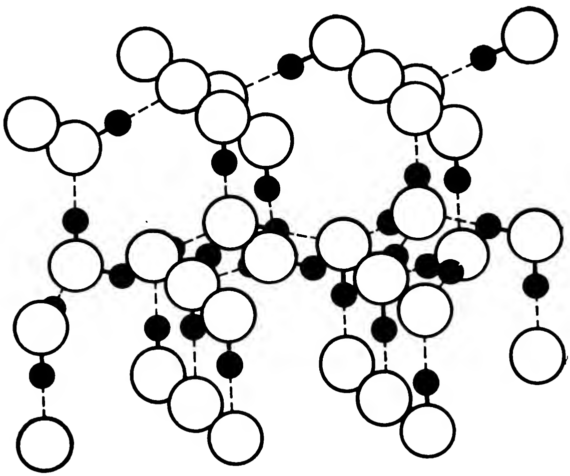


Figure 6

as being the centre of a tetrahedron, with its four closest O neighbours at the apexes. Two solid and two broken lines emerge from the central atom, and the same is true of every atom O, with respect to which a construction identical to the one for the central atom in Figure 6 can be carried out. The white circles represent O atoms; the black, H atoms arranged irregularly, two in the vicinity of any O and one on an O—O line. The randomness consists in that any two of the four lines at each atom can be solid, then the other two are broken. The solid line is shorter, that is, it represents a true chemical bond corresponding to a closer packing of H and O atoms. One connecting line always accommodates one and only one H atom, either closer to the centre or closer to an apex of the tetrahedron. From this it is easy to calculate the number of possible configurations knowing the number of connecting lines.

One mole of ice contains $2N_A$ atoms of hydrogen; each one sits in its tetrahedron in one of two positions on the connecting line. Consequently, a total of 2^4 configurations is possible, neglecting the fact that an oxygen atom must have two hydrogen atoms next

to it to remain a water molecule. The total number of configurations of hydrogen atoms associated with an oxygen atom is 16, namely, four H atoms associated with O (one configuration), three H associated with O (four configurations), two H associated with O (six configurations), one H associated with O (four configurations), no O atoms (one configuration). Of the sixteen configurations, six result in the formation of a molecule of water, that is, for each O we must take 3/8 of the total number of 16 configurations. And since the number of oxygen atoms is N_A, the required number of configurations of water molecules in a crystal of ice is $2^{2N_A}(3/8)^{N_A} = (3/2)^{N_A}$.

This yields the configurational entropy of a mole of ice:

$$S_c = \ln\left(\frac{3}{2}\right)^{N_A} = N_A \ln\frac{3}{2}$$

EXERCISES

1. Find the ratio between the specific heats at constant volume and at constant pressure.

Solution. Using (8.18), we find from the definition of specific heat that

$$c_V = \left(\frac{\partial Q}{\partial \theta}\right)_V = \theta\left(\frac{\partial S}{\partial \theta}\right)_V, \qquad c_p = \left(\frac{\partial Q}{\partial \theta}\right)_p = \theta\left(\frac{\partial S}{\partial \theta}\right)_p$$

The derivatives of implicit functions can be rewritten as follows:

$$\left(\frac{\partial S}{\partial \theta}\right)_V = -\left(\frac{\partial V}{\partial \theta}\right)_S \Big/ \left(\frac{\partial V}{\partial S}\right)_\theta, \qquad \left(\frac{\partial S}{\partial \theta}\right)_p = -\left(\frac{\partial p}{\partial \theta}\right)_S \Big/ \left(\frac{\partial p}{\partial S}\right)_\theta$$

The partial derivatives with the same subscripts can be cancelled out like ordinary fractions since the differentials in them have the same meaning. Then

$$\frac{c_p}{c_V} = \left(\frac{\partial p}{\partial V}\right)_S \Big/ \left(\frac{\partial p}{\partial V}\right)_\theta = \left(\frac{\partial V}{\partial p}\right)_\theta \Big/ \left(\frac{\partial V}{\partial p}\right)_S$$

Thus, specific heats at constant pressure and at constant volume relate in the same way as isothermal compressibility relates to isentropic compressibility. It is sufficient to measure three of the four quantities c_p, c_V, $(\partial V/\partial p)_\theta$, $(\partial V/\partial p)_S$. The fourth can be calculated.

2. Calculate the derivative $(\partial E/\partial V)_\theta$.

Solution. Since $E = F + \theta S$,

$$\left(\frac{\partial E}{\partial V}\right)_\theta = \left(\frac{\partial F}{\partial V}\right)_\theta + \theta\left(\frac{\partial S}{\partial V}\right)_\theta$$

or, taking into account (8.39),

$$\left(\frac{\partial E}{\partial V}\right)_\theta=-p+\theta\left(\frac{\partial p}{\partial \theta}\right)_V$$

If the pressure is known as a function of temperature and volume, the energy can be calculated only up to an arbitrary function of temperature:

$$E=\int dV\left[-p+\theta\left(\frac{\partial p}{\partial \theta}\right)_V\right]+f(\theta)$$

We should therefore always bear in mind that determining the relationship $p = p(V, \theta)$ does not provide complete information concerning the thermodynamic properties of a substance. Furthermore, neither does any pressure term with a linear dependence upon temperature affect the energy, as it is eliminated from the obtained equation. For example, for all ideal gases $p = N\theta/V$, and the energy depends on the temperature in a rather complex way if discrete quantum levels are included in the partition functions.

3. Find $(\partial H/\partial p)_\theta$.

Answer. $V + \theta(\partial V/\partial \theta)_p$.

4. Find the difference between the specific heats at constant pressure and constant volume $(c_p - c_V)$.

Solution. The quantity of heat at constant pressure is dH, and at constant volume, dE (see (8.14) and (8.12)). Thus

$$c_p=\left(\frac{\partial H}{\partial \theta}\right)_p, \qquad c_V=\left(\frac{\partial E}{\partial \theta}\right)_V$$

We transform c_p in the following manner:

$$c_p=\left[\frac{\partial}{\partial \theta}(E+pV)\right]_p=\left(\frac{\partial E}{\partial \theta}\right)_p+p\left(\frac{\partial V}{\partial \theta}\right)_p$$

Further, representing energy as $E = E(\theta, V(p, \theta))$, we write the derivative $(\partial E/\partial \theta)_p$ in the form

$$\left(\frac{\partial E}{\partial \theta}\right)_p=\left(\frac{\partial E}{\partial \theta}\right)_V+\left(\frac{\partial E}{\partial V}\right)_\theta\left(\frac{\partial V}{\partial \theta}\right)_p=c_V+\left(\frac{\partial E}{\partial V}\right)_\theta\left(\frac{\partial V}{\partial \theta}\right)_p$$

whence

$$c_p-c_V=\left(\frac{\partial V}{\partial \theta}\right)_p\left[p+\left(\frac{\partial E}{\partial V}\right)_\theta\right]=\theta\left(\frac{\partial V}{\partial \theta}\right)_p\left(\frac{\partial p}{\partial \theta}\right)_V$$

where we have used the result obtained in Exercise 2. The derivative $(\partial V/\partial \theta)_p$ is transformed thus:

$$\left(\frac{\partial V}{\partial \theta}\right)_p=-\left(\frac{\partial p}{\partial \theta}\right)_V\Big/\left(\frac{\partial p}{\partial V}\right)_\theta$$

whence

$$c_p-c_V=-\theta\left(\frac{\partial p}{\partial \theta}\right)_V^2\Big/\left(\frac{\partial p}{\partial V}\right)_\theta$$

Later (in Sec. 10) it will be rigorously proved that $(\partial p/\partial V)_\theta < 0$, that is, as the volume decreases the pressure can only increase (otherwise the state

of the system will be mechanically unstable). Therefore always $c_p > c_V$ and $(\partial p/\partial V)_S > (\partial p/\partial V)_\theta$ (in accordance with Exercise 1). For ideal gas $(\partial p/\partial\theta)_V = N/V$ and $(\partial p/\partial V)_\theta = -N\theta/V^2$, so that $c_p - c_V = N$.

5. Accepting the second law of thermodynamics as a postulate, prove that the efficiency of a reversible engine is always greater than the efficiency of an irreversible engine operating at the same temperature difference of the source and sink.

Solution. The proof is indirect. Let a reversible engine and an irreversible engine receive the same quantity of heat Q_1, while the quantity of heat Q_2' transferred to the sink from the irreversible engine is less than the quantity of heat Q_2 from the reversible engine. The reversible engine may be made to work as a refrigerator, that is, by applying external work it can be made to transfer heat from a cold reservoir to a hot reservoir. To transfer a quantity of heat Q_1 to the hot reservoir the reversible engine must take from the sink a quantity of heat Q_2 greater than the quantity of heat Q_2' delivered to the sink by the irreversible engine. But in that case it turns out that with the engines operating in opposite directions the hot reservoir receives and delivers the same quantities of heat, that is, it actually does not serve as a heat source. On the other hand, in each cycle the cold reservoir receives a positive quantity of heat $Q_2 - Q_2'$, at the expense of which useful work is performed, the work equal to the difference between the work done by the irreversible engine and the work done by the reversible engine operating as a refrigerator. Thus all the heat is provided by the cold reservoir, which could be simply the surroundings. But this contradicts the second law of thermodynamics.

6. Prove that the specific heat of a system tends to zero when the temperature tends to absolute zero. Prove this also for the derivative $(\partial V/\partial\theta)_p$.

Solution. Entropy is related to specific heat by the equation

$$S = \int_0^\theta \frac{c}{\theta}\, d\theta$$

where the lower limit is set equal to zero in accordance with Nernst's heat theorem. For the integral to exist we must require that $\lim_{\theta\to 0} c = 0$, since otherwise a logarithmic divergence appears. Furthermore, $(\partial V/\partial\theta)_p = -(\partial S/\partial p)_\theta$, and $(\partial S/\partial p)_\theta$ tends to zero when θ tends to zero because $\lim_{\theta\to 0} S = 0$ for an arbitrary value of p.

7. Find the change in temperature of a gas in the Joule-Thomson effect, or $(\partial\theta/\partial p)_H$ as a function of p and θ.

Solution. At constant enthalpy H the required derivative can be written down according to the rule for the differentiation of implicit functions:

$$\left(\frac{\partial\theta}{\partial p}\right)_H = -\left(\frac{\partial H}{\partial p}\right)_\theta \Big/ \left(\frac{\partial H}{\partial\theta}\right)_p.$$

But according to (8.14), the differential of H at constant pressure is equal to the quantity of heat. It is equal, per unit change in temperature, to the specific heat at constant pressure, c_p. In the numerator we express H in terms of the thermodynamic potential from (8.30) and (8.45): $H = G + \theta S$. Then

$$\frac{\partial H}{\partial p} = \left(\frac{\partial G}{\partial p}\right)_\theta + \theta\left(\frac{\partial S}{\partial \theta}\right)_p$$

Using (8.46), we obtain finally

$$\left(\frac{\partial \theta}{\partial p}\right)_H = \frac{1}{c_p}\left[\theta\left(\frac{\partial V}{\partial \theta}\right)_p - V\right]$$

Since for ideal gas $V = N_A\theta/p$, the required quantity is equal to zero.

8. Express the entropy of an ideal gas in terms of the occupation numbers n_k for all three statistics, assuming $g_k = 1$.

Solution. Using the expressions for S from Eqs. (1.14) and (1.24), we find that for

Bose statistics:

$$S = \sum_k [(n_k + 1)\ln(n_k + 1) - n_k \ln n_k]$$

Fermi statistics:

$$S = \sum_k (1 - n_k)\ln(1 - n_k) + n_k \ln n_k$$

Boltzmann statistics for $n_k \ll 1$:

$$S = -\sum_k n_k \ln \frac{n_k}{e}$$

If the weight is not equal to unity, we obtain, respectively,

$$S_{\text{Bose}} = \sum_k g_k [(f_k + 1)\ln(f_k + 1) - f_k \ln f_k]$$

$$S_{\text{Fermi}} = \sum_k g_k [(1 - f_k)\ln(1 - f_k) + f_k \ln f_k]$$

$$S_{\text{Boltzmann}} = -\sum_k g_k f_k \ln \frac{f_k}{e}$$

The first term in S_{Fermi} can be interpreted as the entropy of the "holes", that is, unoccupied levels.

9

THE THERMODYNAMIC PROPERTIES OF IDEAL GAS IN BOLTZMANN STATISTICS

In this section we shall examine some of the corollaries of the general principles of thermodynamics as applied to ideal gases; this will enable a better understanding of the findings of Section 8.

We shall suppose that the gas density is sufficiently small for Boltzmann statistics to be applied to its molecules. This does not mean that the motion of the molecules should be treated as non-quantized: the quantization of rotational, vibrational and, all the more so, electronic levels must be taken into account in all cases when the spacing between neighbouring levels is comparable with, or greater than, θ (that is, with $k_B T$). Even when the spacing of levels is infinitely small compared to θ, as in the case of translational motion, the action quantum should be left in the equation for the statistical weight of the states, otherwise it would be impossible to obtain a unique expression for entropy.

Deviations from Boltzmann statistics that occur in gases at low temperatures or high densities are sometimes termed *degeneracies.* One should distinguish between deviations from the characteristic ideal gas state due to interactions between molecules and quantum deviations from classical statistics. Of course, corrections also arise which are due to the effect of both factors.

Free Energy of an Ideal Gas. As was shown in the preceding section, in calculating thermodynamic quantities it is convenient to proceed from the expression for free energy.

We start with formula (8.40), which must be reduced to the form it takes for a Boltzmann gas. We should keep in mind that a partition function is calculated by definition over all the different states of a gas. But the state of the gas does not change if all possible molecular permutations are performed over the individual states; in non-quantum statistics such a permutation has, in principle, a meaning. The number of permutations of N molecules is equal to $N!$.

The total energy of an ideal gas separates into the sum of the energies of all its molecules:

$$E = \sum_i \varepsilon_i^{(k)}$$

where i is the number of the quantum state. Here E should for the time being be thought of as the actual, not mean, value of the energy.

Substituting the expression for E into the statistical sum (8.40) and dividing this sum by the number of permutations of molecules, we obtain

$$\frac{1}{N!}\sum_k e^{-E/\theta}=\frac{1}{N!}\sum_k \exp\left(-\frac{1}{\theta}\sum_{i=1}^{n}\varepsilon_i^{(k)}\right)$$

$$=\frac{1}{N!}\prod_{i=1}^{N}\sum_k e^{-\varepsilon_i^{(k)}/\theta}=\frac{1}{N!}\left(\sum_k e^{-\varepsilon^{(k)}/\theta}\right)^N \tag{9.1}$$

The second summation over k relates to all possible combinations of the energies of the individual molecules $\varepsilon_i^{(k)}$. Here, we have made use of the fact that the energy spectrum is the same for all the molecules (the gas consists of molecules of the same type). The summation in (9.1) is performed over the spectrum of an individual molecule. Substituting for $N!$ its expression by Stirling's formula, we arrive at a general formula for the free energy of an ideal gas subject to Boltzmann statistics:

$$F=-N\theta\ln\left(\frac{e}{N}\sum_k e^{-\varepsilon^{(k)}/\theta}\right) \tag{9.2}$$

Summation Over the Translational Degrees of Freedom. It is expedient, in the partition function of the states of individual molecules, to separate the translational degrees of freedom and to represent energy in the form

$$\varepsilon=\frac{p^2}{2m}+\varepsilon^{(k)} \tag{9.3}$$

It is assumed here that the gas is not in an external field and the energy cannot therefore depend on the coordinates of the centre of mass of the molecule.

The statistical weight of a state with momentum p is (see Eq. (1.32))

$$g=g^{(k)}\frac{dp_x\,dp_y\,dp_z\,dx\,dy\,dz}{(2\pi\hbar)^3} \tag{9.4}$$

where $g^{(k)}$ is the weight relating to the energy level $\varepsilon^{(k)}$. Integration with respect to x, y, z contributes the factor $\int dx\,dy\,dz=V$ to the partition function. The integration with respect to momenta is performed in the usual way:

$$\int_{-\infty}^{\infty}\exp\left(-\frac{p_x^2}{2m\theta}\right)dp_x=(2\pi m\theta)^{1/2} \tag{9.5}$$

The free energy of an ideal gas thus reduces to

$$F = -N\theta \ln \left[\frac{eV}{N} \left(\frac{m\theta}{2\pi h^2} \right)^{3/2} \sum_k g^{(k)} e^{-\varepsilon^{(k)}/\theta} \right]$$

$$= -N\theta \ln \frac{eVf(\theta)}{N} \tag{9.6}$$

This gives the relationship between free energy and volume. The function $f(\theta)$ depends on the molecular structure.

Thermodynamic Properties of an Ideal Gas. From formula (9.6) it is easy to determine pressure. By (8.39)

$$p = -\left(\frac{\partial F}{\partial V} \right)_\theta = \frac{N\theta}{V} \tag{9.7}$$

This is the well-known ideal gas law.

The thermodynamic potential is

$$G = F + pV = F + N\theta = -N\theta \ln \frac{Vf(\theta)}{N}$$

Here it is expressed in terms of volume. To be able to use the identity (8.45) we must also express V in terms of p, which yields the final formula for the thermodynamic potential G of an ideal gas:

$$G = -N\theta \ln \frac{\theta f(\theta)}{N} \tag{9.8}$$

We find the chemical potential with the help of (8.47) or (8.48):

$$\mu = -\theta \ln \frac{\theta f(\theta)}{N} \tag{9.9}$$

The entropy of an ideal gas is

$$S = -\left(\frac{\partial F}{\partial \theta} \right)_V = N \ln \frac{eVf(\theta)}{N} + N\theta \frac{f'(\theta)}{f(\theta)} \tag{9.10}$$

This expression does not agree with Nernst's heat theorem. Actually, of course, at very low temperatures we must apply to a gas not Boltzmann statistics but quantum statistics, even neglecting the fact that at low temperatures the gas actually condenses.

The energy of the gas is equal to

$$E = F + \theta S$$

or

$$E = N\theta^2 \frac{f'(\theta)}{f(\theta)} = N\theta \frac{d \ln f}{d \ln \theta} \tag{9.11}$$

Thus, the energy of an ideal Boltzmann gas expressed in terms of temperature does not depend on volume at all. The mean energy per molecule, $\bar{\varepsilon} = E/N$, depends only on the temperature of the

gas. This is not only because there is no interaction between gas molecules but also because the properties of the gas are described in terms of classical statistics. In the quantum statistics of ideal gases the energy of a molecule depends on both the volume and the temperature (see Eq. (6.21)). It should be noted that the variables in formula (9.11) do not correspond to the identity (8.26). To make use of this identity one would have to express temperature from (9.10) and substitute it into (9.11), which is difficult to do in general form. The enthalpy of an ideal gas is

$$H = E + pV = N\theta^2 \frac{f'(\theta)}{f(\theta)} + N\theta = N\theta \frac{d \ln \theta f(\theta)}{d \ln \theta} \tag{9.12}$$

Like energy, it depends only on temperature.

A Mixture of Ideal Gases. Since the molecules of ideal gases do not interact, the free energy of a mixture is compounded of the free energies of all its components:

$$F = -\sum_i N_i \theta \ln \frac{eVf_i(\theta)}{N_i} \tag{9.13}$$

The pressure in the mixture is calculated in the usual way:

$$p = -\left(\frac{\partial F}{\partial V}\right)_\theta = \frac{\theta}{V} \sum_i N_i \tag{9.14}$$

If we introduce the *partial pressure* of the *i*th component of the mixture (that is, its contribution to the total pressure p of the gas), then

$$p_i \equiv \frac{N_i \theta}{V} = N_i p \Big/ \sum_n N_n = c_i p \tag{9.15}$$

(see Eq. (8.49)), so that the total pressure appears as the sum of the partial pressures. This refers, of course, only to ideal gases. The thermodynamic potential of the mixture is

$$G = -\sum_i N_i \theta \ln \frac{\theta f_i(\theta)}{p_i} \tag{9.16}$$

The chemical potential of the *i*th component is determined from formula (8.50):

$$\mu_i = \frac{\partial G}{\partial N_i} = -\theta \ln \frac{\theta f_i(\theta)}{p_i} \tag{9.17}$$

These formulas are very important in the theory of chemical equilibria in gases (Sec. 13).

Rotational Energy of a Gas. We shall now calculate the partition function corresponding to the rotational degrees of freedom of molecules. Since we wish to obtain simple formulas, we shall restrict ourselves in this section to the case of nonquantized motion (quantized motion was examined in Sec. 3). This means that the temperature satisfies the condition

$$\theta \gg \frac{h^2}{2I} \tag{9.18}$$

where I is the molecular moment of inertia. At room temperature condition (9.18) is satisfied for all gases including hydrogen.

The expression for the mean energy includes only the logarithmic derivative of the partition function (see Eq. (9.11)), so that the constant factors are of no consequence to the energy. In many applications of statistics, however, the value of the partition function itself is important. To compute this value it must be borne in mind that the summation is taken over physically different molecular states, in other words, over physically different spatial orientations, if motion is treated in the classical sense. For example, the diatomic molecules H_2 or O_2 coincide with themselves in a rotation through 180° around an axis perpendicular to the line joining the nuclei. The position of a diatomic molecule in space is given by two angles, the azimuthal and the polar, and can be represented as a single point on the surface of a sphere of unit radius. But physically different molecular orientations correspond to only one-half of that sphere.

The spatial orientation of a nonlinear molecule with a fixed centre of mass is given with the help of the Euler angles [Sec. 9]. If the molecule possesses any form of symmetry with respect to spatial rotations, the partition function should be divided by the number of ways in which the molecule can be superimposed upon itself in the rotations. For example, the ammonia molecule NH_3 has the shape of a pyramid with a regular triangular base. Its rotational partition function with respect to all spatial orientations should be divided by 3. The benzene molecule C_6H_6 has a regular hexagonal form. A hexagon coincides with itself in a rotation through 60° in its plane and also in a rotation through 180° about an axis joining opposite vertices. Hence, its partition function is taken over $1/(6 \times 2) = 1/12$ of all orientations.

Let us now write the expression for the classical rotational partition function of a diatomic (or, in general, linear) molecule:

$$\sum_{\text{rot}} = \frac{4\pi}{2} \int \int \exp\left(-\frac{M_1^2+M_2^2}{2I\theta}\right) \frac{dM_1\, dM_2}{(2\pi h)^2} \tag{9.19}$$

The factor 4π accounts for all orientations in space. The 2 in the denominator before the integral sign is written for a molecule that

coincides with itself in a rotation through 180° about the axis perpendicular to the line through the nuclei (O_2, CO_2 with the structure O=C=O, etc.). The quantum partition function for an oxygen molecule, whose nuclei have no spin, is taken only with respect to even rotational states (see Sec. 3). This is taken into account in the classical limit by the factor 1/2. In the case of molecular nuclei having spin, the partition function (9.19) must be additionally multiplied by the quantities $(2S+1)$ taken for all the nuclei. Thus, for a linear molecule,

$$\sum_{\text{rot}} = \frac{4\pi}{2} \frac{2\pi I\theta}{(2\pi h)^2} = \frac{2I\theta}{2h^2} \tag{9.20}$$

The position of a nonlinear molecule in space is given by the orientation of an arbitrary axis attached to it and the angle of rotation about that axis. Consequently, all spatial rotations introduce the factor $4\pi \times 2\pi = 8\pi^2$ into the partition function, giving the following expression:

$$\sum_{\text{rot}} = \frac{8\pi^2}{\sigma} \int\int\int \exp\left[-\frac{1}{2\theta}\left(\frac{M_1^2}{I_1}+\frac{M_2^2}{I_2}+\frac{M_3^2}{I_3}\right)\right] \frac{dM_1\,dM_2\,dM_3}{(2\pi h)^3}$$

$$= \frac{(2\theta)^{3/2}\,(\pi I_1 I_2 I_3)^{1/2}}{\sigma h^3} \tag{9.21}$$

Here, the factor σ has the same meaning as the 2 in the denominator of (9.20).

It can be seen from this that the contribution of rotational energy, which is equal to

$$N\theta^2 \frac{d}{d\theta} \ln \sum_{\text{rot}}$$

to the total energy of the gas is $N\theta$ in the case of a linear molecule, and $3N\theta/2$ in the case of a nonlinear molecule.

Vibrational Energy of Molecules. The energy of a molecule performing small oscillations can, according to [7.31], be represented in the following form:

$$\varepsilon_{\text{vib}} = \frac{1}{2} \sum_{\alpha=1}^{n} (P_\alpha^2 + \omega_\alpha^2 Q_\alpha^2) + U_0 \tag{9.22}$$

where Q_α are the normal oscillation coordinates, P_α are the corresponding momenta, and U_0 is the potential energy of the vibrations in the equilibrium state, which was omitted in [7.31]. The classical

limit is attained at $h\omega_\alpha \ll \theta$:

$$\sum_{\text{rot}} = \prod_{\alpha=1}^{n} \int\int \exp\left[-\frac{1}{2\theta}(P_\alpha^2 + \omega_\alpha^2 Q_\alpha^2) - \frac{U_0}{\theta}\right] \frac{dP_\alpha\, dQ_\alpha}{(2\pi h)^n}$$

$$= \frac{(2\pi\theta)^n}{(2\pi h)^n} \prod_{\alpha=1}^{n} \omega_\alpha^{-1} e^{-U_0/\theta} = \left(\frac{\theta}{h}\right)^n \prod_{\alpha=1}^{n} \omega_\alpha^{-1} e^{-U_0/\theta} \tag{9.23}$$

The contribution of vibrations to the mean energy of the gas is $Nn\theta + NU_0$, where n, as is apparent from (9.22), is the number of vibrational degrees of freedom of an individual molecule. Comparing now the results for the mean energy of translational, rotational, and vibrational motions, we see that each quadratic term in the classical expression leads to a contribution of $N\theta/2$ to the total energy of the gas. This is a formulation of the *principle of the equipartition of energy* over the degrees of freedom.

It often happens that $h\omega_\alpha \gg h^2/I$, and there exists a temperature region which satisfies two strong inequalities:

$$\frac{h^2}{I} \ll \theta \ll h\omega_\alpha \tag{9.24}$$

At such temperatures the vibrational quanta of the molecules are still unexcited while the rotational specific heat is already constant. Thus, at temperatures ranging from several tens to several hundred degrees the specific heat of nitrogen and oxygen is $3N/2 + N = 5N/2$. In such conditions gases, notably air, are subject to the equipartition principle over a reduced number of degrees of freedom.

For the equipartition principle to be applicable to vibrational degrees of freedom oscillations with large quantum numbers must be excited. But in that case the molecular states already lie close to the dissociation limit, and some of the molecules separate into atoms. This should be borne in mind in considering the total energy of a gas at high temperatures.

Thermodynamic Properties for a Gas Obeying the Equipartition Principle. The specific heat of a gas obeying the equipartition principle is constant over a wide range of temperatures. Hence, the *ratio of specific heats*

$$\gamma \equiv \frac{c_p}{c_V} = \frac{c_V + N}{c_V} \tag{9.25}$$

is also constant.

It will be convenient, in a number of further applications (Sec. 19), to express thermodynamic quantities in terms of γ. The function $f(\theta)$ is proportional to the quantity

$$\theta^{c_V/N} e^{-U_0/\theta} = \theta^{1/(\gamma-1)} e^{-U_0/\theta}$$

From this we obtain the formula for energy, which we shall write here without the constant term NU_0:

$$E = c_V\theta = \frac{N\theta}{\gamma - 1} = \frac{pV}{\gamma - 1} \tag{9.26}$$

The enthalpy is equal to

$$H = E + pV = \frac{\gamma pV}{\gamma - 1} \tag{9.27}$$

Up to a constant term the entropy is

$$S = N \ln V + \frac{N}{\gamma - 1} \ln \theta = \frac{N}{\gamma - 1} \ln pV^{\gamma} + \text{constant} \tag{9.28}$$

whence we obtain the equation for an isentropic process in a gas obeying the equipartition principle:

$$pV^{\gamma} = \text{constant}$$

The quantity γ is often called the *adiabatic exponent.* Note that in describing real isentropic compression of air we must take $\gamma = 7/5$, which means that a very great compression is needed to excite the vibrational degrees of freedom.

Let us bring together the rules to be used in calculating the specific heat of a gas obeying the equipartition principle. For each translational and rotational degree of freedom we must put $N/2$, since they yield one quadratic variable each in the Hamiltonian of a molecule. Hence the classical partition function acquires either the factor $(2\pi m\theta)^{1/2}$ or the factor $(2\pi I\theta)^{1/2}$. If a molecule is linear, there are two rotational degrees of freedom, and three if it is nonlinear.

A molecule consisting of i atoms has $3i$ degrees of freedom. Hence either $3i - 5$ or $3i - 6$ of the degrees of freedom are involved in vibrations. Each makes a contribution N to the specific heat if $h\omega \ll \theta$.

Thus, for a triatomic molecule of triangular configuration, for example H_2O, full excitation of all degrees of freedom (other than electronic) yields a specific heat of $c_V = 6N$, $\gamma = 7/6$. If vibrations have not yet been excited, then $c_V = 3N$, $\gamma = 4/3$. At the lowest temperature only the translational degrees of freedom remain, as in the case of a monatomic gas, which yields $c_V = 3N/2$, $\gamma = 5/3$.

If the atoms of a triatomic molecule form a straight line (as is the case for CO_2), the maximum specific heat $c_V = 13/2$, $\gamma = 15/13$, that is, c_V is greater for a linear molecule than for a triangular molecule. But if vibrations are not excited, then $c_V = 5/2$, which is less than for a triangular molecule. Such an intersection of the specific heat curves of CO_2 and H_2O when temperature changes is actually observed.

Polymeric Chains. There exists a state of condensed bodies, called *polymers*, which in some respects resembles the gaseous state—if not in the strict, quantitative, sense then at least qualitatively. There are common features in the behaviour of gases and substances made up of very long and flexible polymeric molecules, such as rubber. Its molecules are made up of separate *monomers* coupled in an extremely flexible fashion.

Ordinarily the angle between neighbouring monomers is fairly well defined due to the directional nature of valence forces. But for a given angle it is possible for one monomer to revolve almost freely about an axis passing through a neighbouring monomer. In more precise terms, the interaction energy between monomers depends weakly on this spatial rotation, so that at room temperature a free, random rotation takes place resembling the motion of unbonded gas molecules.

Random rotation of the monomers in space results in the molecules getting tangled into a ball. Obviously, in such a state its entropy is greater than when it is drawn out, since the number of coiled configurations is considerably greater than the number of extended configurations. A molecule coiled into a ball possesses greater configurational entropy (see Sec. 8).

A real polymer consists of monomers of finite volume. In a detailed study of its states one would have to consider the reciprocal impenetrability of the monomers. That is an extremely difficult task. But for a qualitative understanding of the thermodynamic properties of polymers it is sometimes sufficient to treat the monomers as freely intersecting and hinged together in a way allowing complete freedom of rotation over the whole surface of a sphere whose centre is at the hinge point.

Such a picture is favoured by the following consideration. In the isothermal reversible stretching of rubber its total energy remains unchanged: the amount of heat transmitted to the surroundings is dQ, taken with the sign opposite to that of the work dA done in the stretching. But $dA = \tau\, dl$, where dl is the elongation, and τ is the tension. The plus sign has been chosen so that positive work is done on the polymer in increasing its length.

Since in a reversible process $dQ = \theta\, dS$, we find that $\tau\, dl = -\theta\, dS$. The elongation of rubber is associated with a decrease in its entropy, as stated. From the latter relationship we obtain

$$\tau = -\theta \left(\frac{\partial S}{\partial l}\right)_E \tag{9.29}$$

By analogy, in a gas, from (8.16),

$$p = \theta \left(\frac{\partial S}{\partial V}\right)_E$$

If from data on specific heat at various elongations we know the entropy as a function of length, that is, the distance between the ends of a thread in a ball, from Eq. (9.29) we can calculate the tension as a function of length.

Entropy of a Chain With Free Rotation. We shall now calculate the entropy of a chain with freely rotating couplings. The assumption of spatial rotation at a constant valence angle between monomers would be closer to reality, but the general nature of the dependence of entropy upon l is apparently the same in such a model as in the one with free rotation.

We shall assume that a chain consists of N monomers. Let us determine the probability that these monomers will, upon forming a polymeric chain, yield a distance l between its ends. In a greatly tangled coil $l \ll Nb$, where b is the length of one monomer.

As usual, probability is determined by the number of modes in which the chain can be arranged in space between the given initial and terminal points. The logarithm of the probability is the configurational entropy (according to Sec. 8).

Denote the vector between the ends of the chain by $\mathbf{l}$, and the vector of an individual monomer by $\mathbf{b}_k$. We assume that these vectors have the same length b but different spatial orientations. Then the probability density of each individual spatial configuration of the whole chain is equal to

$$\delta\left(\mathbf{l}-\sum_k \mathbf{b}_k\right)$$

Here the δ function is given by [26.28]. It is equal to zero for all the positions of the vectors $\mathbf{b}_k$ except those that in sum yield vector $\mathbf{l}$. In turn, the integral of that function over all the values of $\mathbf{l}$ is 1. The required probability that the vector between the ends of the chain is equal to $\mathbf{l}$ is obtained if the δ function is averaged over all the spatial configurations of the individual monomers:

$$W_N(\mathbf{l})=\frac{1}{(4\pi)^N}\int d\Omega_1 \ldots \int d\Omega_N\ \delta\left(\mathbf{l}-\sum_{k=1}^{N}\mathbf{b}_k\right) \tag{9.30}$$

To perform the averaging, it is convenient to represent the δ function in terms of a Fourier integral:

$$\delta\left(\mathbf{l}-\sum_{k=1}^{N}\mathbf{b}_k\right)=\frac{1}{8\pi^3}\int dV_\xi \exp\left[i\xi\left(\mathbf{l}-\sum_{k=1}^{N}\mathbf{b}_k\right)\right]$$

This is obtained from the general definition [26.28] by substituting into it the eigenfunctions of the momentum operator $(2\pi h)^{-3/2}e^{i\mathbf{pr}/h}$ with $h=1$. Now we can interchange the order of integration over

V_ξ and Ω_k. Each such averaging yields

$$\frac{1}{4\pi}\int e^{i\xi \mathbf{b}_k}\,d\Omega_k = \frac{\sin b\xi}{b\xi} \tag{9.31}$$

After this W_N (l) reduces to the form

$$W_N(\mathbf{l}) = \frac{1}{8\pi^3}\int e^{i\xi\mathbf{l}}\left(\frac{\sin b\xi}{b\xi}\right)^N dV_\xi \tag{9.32}$$

We now make use of the fact that N is large. Then only small values of ξ, for which $(\sin b\xi)/(b\xi)$ is close to unity, can make an appreciable contribution to the integral.

Accordingly, we must take

$$\left(\frac{\sin b\xi}{b\xi}\right)^N = \left(1-\frac{b^2\xi^2}{6}\right)^N$$

The subsequent terms of the sine expansion make a contribution that tends to zero with respect to the basic result as N increases.

The length of an extended chain is $Nb \equiv L$. We now perform the following limiting transition:

$$\left(1-\frac{b^2\xi^2}{6}\right)^L = \left(1-\frac{Lb\xi^2}{6N}\right)^N_{N\to\infty} \to e^{-Lb\xi^2/6}$$

Then, after averaging over the directions ξ, we obtain

$$W_L(l) = \frac{1}{4\pi^2 li}\int_0^\infty (e^{i\xi l} - e^{-i\xi l})\,e^{-Lb\xi^2/6}\xi\,d\xi \tag{9.33}$$

For the present it is convenient to leave this expression in its complex form. In the second term we replace ξ by $-\xi$, and we obtain the following integral:

$$W_L(l) = \frac{1}{4\pi^2 li}\int_{-\infty}^{\infty} e^{i\xi l - Lb\xi^2/6}\xi\,d\xi$$

We represent the exponent in the integrand as follows:

$$-\left(\xi^2\frac{bL}{6} - i\xi l\right) = -\frac{bL}{6}\left(\xi^2 - \frac{6i\xi l}{bL} - \frac{9l^2}{b^2L^2}\right) - \frac{3}{2}\frac{l^2}{bL}$$
$$= -\frac{bL}{6}\left(\xi - \frac{3il}{bL}\right)^2 - \frac{3}{2}\frac{l^2}{bL}$$

Going over to a new variable $\xi' = \xi - 3il/(bL)$ and making use of the fact that the integral of an odd function is zero, we obtain

$$W_L(l) = \left(\frac{3}{2\pi bL}\right)^{3/2}\exp\left(-\frac{3}{2}\frac{l^2}{bL}\right) \tag{9.34}$$

Using the formulas of Exercise 3 in Section 1, we can verify that this distribution is normalized to unity:

$$4\pi \int_0^{\infty} l^2 W_L(l)\, dl = 1$$

Thus, up to a term that does not depend on l, the entropy of the whole chain for given L and l is

$$S = \ln W_L(l) = -\frac{3}{2}\frac{l^2}{bL} \tag{9.35}$$

By (9.29) the tension in the chain is

$$\tau = -\theta \frac{\partial S}{\partial l} = \frac{3l\theta}{bL} \tag{9.36}$$

Thus, the tension is proportional to the length of the chain. This resembles Hooke's law. Like gas pressure p, tension τ is "entropic" in character. In nonpolymeric condensed bodies, tension τ is primarily due to the interactions between molecules.

EXERCISES

1. Determine the work done on, and quantity of heat received by, a gas in an isothermal process.

Solution. The work equals the change in free energy:

$$A = -N\theta \ln (V_2/V_1)$$

The quantity of heat is expressed in terms of the change in entropy:

$$Q = \theta (S_2 - S_1) = N\theta \ln (V_2/V_1)$$

Quantities A and Q are equal in magnitude and opposite in sign because at constant temperature the energy of an ideal gas does not change.

2. Two portions of different gases having the same temperature and pressure are mixed. Determine the increase in entropy.

Solution. Using the known expression for entropy, we can write

$$\Delta S = N_1 \ln \frac{\theta f_1(\theta)}{p_1} + N_2 \ln \frac{\theta f_2(\theta)}{p_2} - N_1 \ln \frac{\theta f_1(\theta)}{p_1} - N_2 \ln \frac{\theta f_2(\theta)}{p_2}$$

where p_1 and p_2 are the partial pressures of both gases after mixing. Hence

$$\Delta S = N_1 \ln \frac{p}{p_1} + N_2 \ln \frac{p}{p_2} = N_1 \ln \frac{N_1 + N_2}{N_1} + N_2 \ln \frac{N_1 + N_2}{N_2}$$

If two portions of the same gas are mixed under the same conditions, the entropy after mixing is equal to $(N_1 + N_2) \ln (\theta f_1/p)$, and $\Delta S = 0$, as it should be. This result could not have been obtained without multiplying the partition function by the coefficient $(N!)^{-1}$. Due to this factor, only the summation over physically different states of the gas is involved in the free energy, and the entropy does not change when two portions of the same gas of equal temperature and pressure are brought together. Note that when formulas of Boltzmann statistics are developed from quantum statistics by means of a limiting transition, the required factor is obtained automatically.

3. Calculate the free energy of a gas in a centrifuge of radius R and length l rotating at an angular velocity ω. Find the mean square distance of a molecule from the axis.

Solution. The centrifugal force is equal to $m\omega^2 r$ (see [8.8]), which corresponds to an effective potential energy $U = -\int m\omega^2 r\, dr = -m\omega^2 r^2/2$, whence we obtain the expression for the free energy:

$$F = -N\theta \ln \left[\frac{ef(\theta)}{N} l \int_0^R \exp\left(\frac{m\omega^2 r^2}{2\theta}\right) r\, dr\right]$$

$$= -N\theta \ln \left\{\frac{ef(\theta)}{N} \frac{\theta l}{m\omega^2} \left[\exp\left(\frac{m\omega^2 R^2}{2\theta}\right) - 1\right]\right\}$$

The free energy satisfies the general relation $dF = -S\, d\theta - \Lambda\, d\lambda$, where ω^2 must be regarded as an external parameter (see Eq. (8.38)).

We now determine the mean square distance of a molecule from the axis,

$$\overline{r^2} = -\frac{2}{Nm} \frac{\partial F}{\partial(\omega^2)} = \left\{\left[1 - \exp\left(-\frac{m\omega^2 R^2}{2\theta}\right)\right]^{-1} - \frac{2\theta}{m\omega^2 R^2}\right\} R^2$$

because, if $\omega^2 \equiv \lambda$, then $m\overline{r^2}/2 = -\partial E/\partial(\omega^2) = -\partial F/\partial\lambda \equiv \Lambda$ (see (8.7)).

At very large angular velocities $\overline{r^2} \to R^2$, and at small velocities $\overline{r^2} = R^2/2$.

10

FLUCTUATIONS

The Reversibility of Mechanical Equations with Respect to Time. In classical mechanics the initial and final states of a system are uniquely related: one completely determines the other. Mathematically, this is expressed by the fact that if the signs of all the velo-

cities are reversed, the motion will occur in the reverse direction. Changing the sign of the velocities is formally equivalent to changing the sign of time. But the form of the Lagrange equations does not change when changing the sign of time. The same is true of Newton's Second Law, which involves only second derivatives with respect to time.

In order to perform the same transition in the equations of electrodynamics we must first change the signs of the currents. For the form of Maxwell's equations [12.30] and [12.32] to remain unchanged we must change the sign of the magnetic field together with that of the current, leaving the electric field unchanged. Since the vector of magnetic field is axial, or a pseudovector, the choice of sign is purely a question of convenience.

Quantum mechanical equations also preserve their form when t is changed to $-t$. In the simplest case, when the Hamiltonian operator is real, that is, does not involve i, the transition from t to $-t$ means simply the transition $\psi \rightarrow \psi^*$, which can be directly seen from the Schrödinger equation $\hat{\mathscr{H}}\psi = ih\,(\partial\psi/\partial t)$. But the function ψ is completely equivalent to ψ^* (it does not matter which of them is regarded as conjugate). In more complicated cases, when $\hat{\mathscr{H}}$ is complex, we can likewise always pass, when t changes to $-t$, from the function ψ to another that is physically equivalent to it.

Statistical Mechanics and the Reversibility of Time. We shall now investigate how the laws of statistical mechanics relate to time reversal. Statistical mechanics states that if at some initial moment of time a system deviates from statistical equilibrium, then in most cases it will subsequently approach equilibrium. A system in equilibrium under unchanging external conditions will remain in equilibrium whatever imaginable changes in the sign of time are made in the equations of mechanics describing the detailed microscopic state of the system. A situation therefore arises which at first sight seems paradoxical: statistical laws, which appear noninvariant with respect to time reversal, are deduced from the equations of mechanics.

Statement of the Problem in Classical Mechanics and Statistics. The contradiction disappears at once if we deal with quasi-closed statistical systems interacting with the surroundings. In such systems the replacement of t by $-t$ is simply not enough to make processes proceed in the reverse direction. It is of interest, however, to examine the situation in the case of a strictly closed system so as to bring closer the general statement of the problem in mechanics and statistics. Let us examine an apparent paradox within the limits of the classical laws of motion. First take the following example. Let a gas

occupy one half of a vessel divided by a partition, and let the other half be evacuated. When the partition is removed, the gas fills the whole vessel. We consider the motion of each molecule as obeying the laws of classical mechanics, so that at each moment of time we know exactly its coordinates and momenta. Let them be plotted on the axes of a $6N$-dimensional coordinate system encompassing the *phase space*. Every point in this space gives the state of all the molecules of the gas. The motions of all the molecules will be represented in the form of the motion of a single point along a path in the phase space.

Thus, the transition of the gas from the state in which it was assembled completely in one half of the vessel to a state of statistical equilibrium corresponds to the displacement of a point in phase space from one domain to another domain corresponding to statistical equilibrium. If the gas is completely isolated from external influences and if in the state of statistical equilibrium the signs of all the velocities are by some device reversed, the phase point representing the state of the gas will start moving in the opposite direction, and all the gas will gather in one half of the vessel. And since any equilibrium state of a gas is attainable from a nonequilibrium state, it would seem that the gas should come out of the state of statistical equilibrium just as often as it enters it. Actually this, of course, is never observed. That is, in its simplest form the apparent contradiction between the classical mechanical principle of causality (and the reversibility of time associated with it) and the concept of the irreversibility of transitions in statistics.

In actual fact, in statistics equilibrium is not just one strictly defined state but a whole range of states in which a closed system spends most of its time. The phase point moves about the equilibrium domain for an extremely long time before spontaneously leaving it to any appreciable distance. Through the vast majority of phase points in the statistical equilibrium domain there pass such paths whose "convolutions" almost never enter domains corresponding to appreciable nonequilibrium states.

If we choose a certain volume of the equilibrium domain, we can say that the system leaves it just as often as it enters it but that in the overwhelming majority of cases it does not go "far" from this volume.

Therefore, the apparent irreversibility in statistics is associated with the way the problem is stated. Namely, a system does not remain for long in nonequilibrium states and quickly enters equilibrium states; it remains in equilibrium states for a very long time, so that the probability of spontaneously leaving these states can for all practical purposes be neglected. In this section, however, we shall calculate the probability of spontaneous, albeit small, deviations of a system from equilibrium.

Quantum Mechanics and the Irreversibility of Transitions. Fundamental to quantum statistics is the principle of detailed balance (see Sec. 1). In accordance with this principle, the probabilities of direct and reverse transitions between two states having the same statistical weights are equal. However, it by no means follows from this principle that the probability of a transition from an equilibrium state to a nonequilibrium state is the same as that of a transition from a nonequilibrium state to an equilibrium state. A state of statistical equilibrium includes a great number of equiprobable microstates, while a nonequilibrium state includes comparatively few microstates: a system spends most of the time in equilibrium for the very reason that the number of nonequilibrium states is incomparably smaller than the number of equilibrium states. Every given microstate belonging to a set of statistical equilibrium states passes to a state in the same domain (that is, an equilibrium state) with overwhelming probability, while the probability of a transition into a nonequilibrium state is negligible. A nonequilibrium state passes preferentially into an equilibrium state because a transition to a state of less equilibrium can occur in an incomparably smaller number of equiprobable ways. That is why a system tends to equilibrium despite the equal probability of direct and reverse transitions between any two equiprobable microstates.

Poisson Distribution Formula. A spontaneous transition of a system from a state of equilibrium to an appreciably nonequilibrium state is highly improbable, but not altogether impossible. The deviations of actual values from the mean are the more probable the smaller the system in which they occur. If, for example, gas molecules are observed in a cube with a side 10^{-6} cm, under normal conditions (0 °C, 760 mmHg) the mean number of molecules is only 27. Molecules may pass into neighbouring sections, so that the actual number in a certain volume will exhibit very noticeable deviations from the number 27.

It is easy to determine the probability that N molecules will occur in a given volume V if the total volume V_0 contains N_0 molecules. The probability of one molecule occuring in the volume V is, obviously, V/V_0. Hence, the probability of N molecules occurring in the volume V and of $(N_0 - N)$ molecules occurring in the remaining part of the volume, is equal to

$$\frac{N_0!}{(N_0-N)!\,N!}\left(\frac{V}{V_0}\right)^N\left(1-\frac{V}{V_0}\right)^{N_0-N} \tag{10.1}$$

The first factor indicates the number of ways in which N molecules can be selected out of the total number N_0. A formula analogous to (10.1) was derived in Exercise 1, Section 1, for the probability of tails occurring N times.

Let the total number of molecules, N_0, be arbitrarily great, and let N be much smaller than N_0. We transform the ratio of factorials as follows:

$$\frac{N_0!}{(N_0-N)!}=N_0(N_0-1)(N_0-2)\ldots(N_0-N+1)$$
$$=N_0^N\left(1-\frac{1}{N_0}\right)\left(1-\frac{2}{N_0}\right)\ldots\left(1-\frac{N-1}{N_0}\right)=N_0^N$$

We represent the quantities $(1-V/V_0)^{N_0-N}$ and $(V/V_0)^N$ as follows:

$$\left(1-\frac{V}{V_0}\right)^{N_0-N}=\left[\left(1-\frac{\overline{N}}{N_0}\right)^{N_0}\right]^{(1-N/N_0)}$$
$$\approx(e^{-\overline{N}})^{(1-N/N_0)}\approx e^{-\overline{N}}$$
$$\left(\frac{V}{V_0}\right)^N=\frac{\overline{N}^N}{N_0^N}$$

where, by definition of $\overline{N}$, $\overline{N}=N_0\,(V/V_0)$. Substituting all the obtained expressions into the initial formula, we obtain the required probability

$$W_N=e^{-\overline{N}}\frac{(\overline{N})^N}{N!} \tag{10.2}$$

This is the *Poisson distribution.* It can be seen at once that, like the initial expression, it is normalized to unity. It will be shown in Exercise 1 that the distribution (10.2) has a very sharp maximum at $N=\overline{N}$.

A Poisson distribution is used to calculate the fluctuations of random quantities in the most diverse cases, for instance, for the number of counts in the counter of ionizing particles. It is used to determine the probability that the counts indicate a real effect and are not due to noise, which allows for a certain excess count over the mean.

Fluctuation Probability. Let us deduce a general formula for the probability of fluctuation of quantities in a subsystem of a large system. The small volume of gas just considered is a special case of such a subsystem.

Let a spontaneous deviation from statistical equilibrium have occurred in a subsystem. Thereby, the whole system will have deviated from equilibrium. The ratio of the probabilities of the equilibrium and nonequilibrium states of a large system is equal to the ratio of the statistical weights of the states:

$$\frac{W}{W_0}=\frac{G}{G_0} \tag{10.3}$$

where W and G refer to the large system. The subscript 0 denotes the equilibrium state.

Expressing the statistical weight in terms of entropy ($S = \ln G$), we obtain

$$\frac{W}{W_0} = e^{S-S_0} \tag{10.4}$$

Equation (10.4) can be given a somewhat different form. Since the large system is closed, its energy does not change in a fluctuation ($E = E_0$). But the relationship between the total and free energies is: $F = E - \theta S$, $F_0 = E_0 - \theta S_0$. It follows from these equations that the change in entropy of a system undergoing a fluctuation is equal to the change in free energy, taken with opposite sign and divided by the temperature:

$$S - S_0 = \frac{F_0 - F}{\theta} = -\frac{A_{\min}}{\theta} \tag{10.5}$$

The change in free energy is expressed in terms of minimum work with the help of (8.36).

We note at once that $A_{\min}$ is the minimum *external* work that would have to be done of the system to produce the given fluctuation reversibly, that is, without any change in entropy. Actually no work is done, the fluctuation is spontaneous, and a certain decrease in entropy occurs. That work is not in fact done is seen from the equation $E = E_0$, on which the derivation was based.

Thus, the fluctuation probability of a subsystem is given by the following formula derived by Einstein:

$$W \sim e^{-A_{\min}/\theta} \tag{10.6}$$

The same deviation as indicated by formula (10.6) can be caused by performing reversible work $A_{\min}$.

Fluctuations of Thermodynamic Quantities. Let us reduce the expression for minimum energy to a form more convenient for actual calculations. We assume that a large system is divided into two parts: a smaller one, in which the fluctuation occurs, and the larger remaining part, in which the variation of the quantities is reversible. *Fluctuation* is a spontaneous disturbance of statistical equilibrium in a small subsystem. We shall write the fluctuating quantities relating to the subsystem without subscripts, and the equilibrium quantities with the subscript zero; quantities describing the remaining part of the system are primed.

By definition, the minimum work is calculated in the case of constant entropy of the whole system, that is, as though instead of a fluctuation occurring there is a certain change in quantities due to an external action that does not destroy the statistical equilibrium.

In an external action the work is equal to the change in energy of the system:

$$A_{\min} = \Delta E + \Delta E' \tag{10.7}$$

The positive sign corresponds to the fact that work is done on the system.

The changes in the quantities of the large system are very small, and they are the smaller the larger the system, so that $\Delta E'$ can be replaced according to the thermodynamic identity (8.26):

$$\Delta E' = \theta_0 \Delta S' - p_0 \Delta V' \tag{10.8}$$

The work $A_{\min}$ is calculated, as pointed out before, in the case of a reversible process. Consequently, $\Delta S' = -\Delta S$, and besides, of course, $\Delta V' = -\Delta V$. Hence

$$A_{\min} = \Delta E - \theta_0 \Delta S + p_0 \Delta V \tag{10.9}$$

Large fluctuations are highly improbable, therefore the quantities ΔS and ΔV should be regarded as small in the subsystem as well. But for a subsystem it is already necessary to make a series expansion up to second order quantities, since otherwise the expression for $A_{\min}$ would be identically zero (close to the maximum the entropy expansion can begin only with quadratic terms). Thus

$$\Delta E = \left(\frac{\partial E}{\partial S}\right)_0 \Delta S + \left(\frac{\partial E}{\partial V}\right)_0 \Delta V + \frac{1}{2}\left(\frac{\partial^2 E}{\partial V^2}\right)_0 (\Delta V)^2 + \left(\frac{\partial^2 E}{\partial V\, \partial S}\right)_0 \Delta V\, \Delta S + \frac{1}{2}\left(\frac{\partial^2 E}{\partial S^2}\right)_0 (\Delta S)^2$$

But $(\partial E/\partial S)_0 = \theta_0$ and $(\partial E/\partial V)_0 = -p_0$, which implies that only quadratic terms in small deviations remain in the expression $A_{\min} = \Delta E + \Delta E'$. Taking advantage of the fact that

$$\left(\frac{\partial^2 E}{\partial V^2}\right)_S = -\left(\frac{\partial p}{\partial V}\right)_S, \qquad \left(\frac{\partial^2 E}{\partial S^2}\right)_V = \left(\frac{\partial \theta}{\partial S}\right)_V$$

$$\frac{\partial^2 E}{\partial S\, \partial V} = -\left(\frac{\partial p}{\partial S}\right)_V = \left(\frac{\partial \theta}{\partial V}\right)_S$$

we write $A_{\min}$ as follows:

$$A_{\min} = -\frac{1}{2}\Delta V\left[\left(\frac{\partial p}{\partial V}\right)_0 \Delta V + \left(\frac{\partial p}{\partial S}\right)_0 \Delta S\right] + \frac{1}{2}\Delta S\left[\left(\frac{\partial \theta}{\partial V}\right)_0 \Delta V + \left(\frac{\partial \theta}{\partial S}\right)_0 \Delta S\right]$$
$$= \frac{1}{2}(\Delta\theta\, \Delta S - \Delta p\, \Delta V) \tag{10.10}$$

Therefore, (10.6) transforms to the form

$$W \sim \exp\left[\frac{1}{2\theta}(\Delta p\, \Delta V - \Delta\theta\, \Delta S)\right] \tag{10.11}$$

where the subscript 0 is omitted from θ.

Let us find the probability for the volume and temperature fluctuations. For this replace Δp and ΔS with their expressions in terms of volume and temperature:

$$\Delta p = \left(\frac{\partial p}{\partial V}\right)_\theta \Delta V - \left(\frac{\partial p}{\partial \theta}\right)_V \Delta\theta$$

$$\Delta S = \left(\frac{\partial S}{\partial V}\right)_\theta \Delta V + \left(\frac{\partial S}{\partial \theta}\right)_V \Delta\theta$$

But $(\partial p/\partial\theta)_V = (\partial S/\partial V)_\theta$ by (8.39); hence the right-hand side of Eq. (10.11) can be represented as the product of two factors, one dependent on ΔV and the other on $\Delta\theta$:

$$W = \exp\left[\frac{1}{2\theta}\left(\frac{\partial p}{\partial V}\right)_\theta (\Delta V)^2\right] \exp\left[-\frac{1}{2\theta}\left(\frac{\partial S}{\partial \theta}\right)_V (\Delta\theta)^2\right] \tag{10.12}$$

It is now easy to determine the mean square fluctuations $\overline{(\Delta V)^2}$ and $\overline{(\Delta\theta)^2}$. Introducing the notation

$$\frac{1}{2\theta}\left(\frac{\partial p}{\partial V}\right)_\theta = -\alpha \tag{10.13}$$

we can write the square of volume fluctuation as follows:

$$(\Delta V)^2 = -\frac{\partial}{\partial\alpha} \ln \int_{-\infty}^{\infty} e^{-\alpha(\Delta V)^2} d(\Delta V) = -\frac{\partial}{\partial\alpha} \ln\left(\frac{\pi}{\alpha}\right)^{1/2} = \frac{1}{2\alpha} \tag{10.14}$$

The integration was legitimately extended over the whole numerical axis from $-\infty$ to ∞, since at large ΔV the integrand is very small. We finally arrive at the formula

$$\overline{(\Delta V)^2} = -\theta \Big/ \left(\frac{\partial p}{\partial V}\right)_\theta \tag{10.15}$$

It should be noted that this expression for the volume fluctuation of a subsystem is applicable only for the case of constant temperature. At constant entropy the expression would have been different. Similarly we find the mean square of the temperature fluctuation at constant volume:

$$\overline{(\Delta\theta)^2} = \theta\left(\frac{\partial\theta}{\partial S}\right)_V = \frac{\theta^2}{c_V} \tag{10.16}$$

Note that the square of the fluctuation in volume (an additive quantity) is proportional to the first power of the additive quantity $(\partial/V\partial p)_\theta$. Hence, the relative volume fluctuation $\sqrt{\overline{(\Delta V)^2}/V}$ is inversely proportional to the square root of the dimensions of the system. This statement as applied to energy was made in Section 7. The temperature fluctuation $\sqrt{\overline{(\Delta\theta)^2}}$ is inversely proportional to the

square root of the specific heat, and therefore also decreases with the dimensions of the subsystem, as could be expected.

The quantity θ is the modulus of the Gibbs distribution for the entire large system. When a fluctuation occurs in a subsystem, the quantity θ does not, naturally, coincide with its temperature, that is, with the distribution modulus that refers to the time interval during which the subsystem was independent of the larger system. Strictly speaking, during that time interval θ was not the temperature of the large system either, since θ has the meaning of temperature only in total equilibrium.

The temperature and energy of a system are related not completely unambiguously: at a given energy the temperature can experience small fluctuations, and at a given temperature energy may fluctuate.

Thermodynamic Inequalities. Two very important thermodynamic inequalities follow from formulas (10.15) and (10.16):

$$\left(\frac{\partial p}{\partial V}\right)_\theta < 0, \qquad c_V > 0 \tag{10.17}$$

The state of a substance can be stable only when these inequalities are satisfied. If this is not the case, the deviation from equilibrium becomes the more probable the greater it is. If the equation of state of a substance indicates that the inequalities (10.17) break down at some values of p, θ, or V, the substance is unstable in that domain of values of the thermodynamic quantities, and it must break up into separate phases (for example, liquid or vapour), to which other values of V correspond.

The Mean Value of the Fluctuations of Two Quantities. Let us now consider the fluctuations of volume and entropy together. In this case the fluctuation probability formula has the form

$$W \sim \exp\left\{-\frac{1}{2\theta}\left[-\left(\frac{\partial p}{\partial V}\right)_S (\Delta V)^2 + 2\left(\frac{\partial \theta}{\partial V}\right)_S \Delta V\,\Delta S + \left(\frac{\partial \theta}{\partial S}\right)_V (\Delta S)^2\right]\right\} \tag{10.18}$$

Here the expression on the right-hand side no longer separates into a product of two factors that depend on each variable separately. Therefore, like the volume fluctuation at constant entropy and the entropy fluctuation at constant volume, the mean value of the product of the volume and entropy fluctuations also differs from zero: $\overline{\Delta V\,\Delta S} \neq 0$. Let us calculate this mean value according to formula (10.18), which we first write in abbreviated notation:

$$W \sim f(\alpha_{11}, \alpha_{12}, \alpha_{22}) \equiv \exp\left\{-\frac{1}{2}\left[\alpha_{11}(\Delta V)^2 + 2\alpha_{12}\Delta V\,\Delta S + \alpha_{22}(\Delta S)^2\right]\right\} \tag{10.19}$$

The required quantity is then

$$\overline{\Delta V\,\Delta S} = -\frac{\partial}{\partial \alpha_{22}} \ln \int\int_{-\infty}^{\infty} f(\alpha_{11}, \alpha_{12}, \alpha_{22})\, d(\Delta V)\, d(\Delta S)$$

To calculate the integral we write the quadratic expression in the exponent in the form of a sum of quadratic terms:

$$\alpha_{11}\left(\Delta V + \frac{\alpha_{12}}{\alpha_{11}}\Delta S\right)^2 + \frac{\alpha_{11}\alpha_{22} - \alpha_{12}^2}{\alpha_{11}} (\Delta S)^2$$

Now we change variables:

$$\Delta V + \frac{\alpha_{12}}{\alpha_{11}} \Delta S \equiv x$$

The integration variable x varies within the same limits as ΔV and ΔS, that is, from $-\infty$ to ∞. The integral in (10.19) can now be calculated as follows:

$$\int\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left[\alpha_{11}x^2 + \frac{\alpha_{11}\alpha_{22} - \alpha_{12}^2}{\alpha_{11}} (\Delta S)^2\right]\right\} dx\, d(\Delta S)$$

$$= 2\left(\frac{\pi}{\alpha_{11}}\right)^{1/2} \left(\frac{\pi\alpha_{11}}{\alpha_{11}\alpha_{22} - \alpha_{12}^2}\right)^{1/2} = \frac{2\pi}{(\alpha_{11}\alpha_{22} - \alpha_{12}^2)^{1/2}} \tag{10.20}$$

whence

$$\overline{\Delta V\,\Delta S} = -\frac{\partial}{\partial \alpha_{12}} \ln \frac{2\pi}{(\alpha_{11}\alpha_{22} - \alpha_{12}^2)^{1/2}}$$

$$= \frac{\alpha_{12}}{\alpha_{12}^2 - \alpha_{11}\alpha_{22}} = \frac{\theta\left(\frac{\partial\theta}{\partial V}\right)_S}{\left(\frac{\partial\theta}{\partial V}\right)_S^2 - \left(\frac{\partial p}{\partial V}\right)_S \left(\frac{\partial\theta}{\partial S}\right)_V} \tag{10.21}$$

Consider the quantity

$$\frac{\theta}{\overline{\Delta V\,\Delta S}} = \left(\frac{\partial\theta}{\partial V}\right)_S^2 + \left(\frac{\partial p}{\partial V}\right)_S \left(\frac{\partial\theta}{\partial S}\right)_V \Big/ \left(\frac{\partial\theta}{\partial V}\right)_S$$

$$= \left(\frac{\partial\theta}{\partial V}\right)_S - \left(\frac{\partial p}{\partial V}\right)_S \left(\frac{\partial V}{\partial S}\right)_\theta$$

$$= -\left(\frac{\partial p}{\partial S}\right)_V - \left(\frac{\partial p}{\partial V}\right)_S \left(\frac{\partial V}{\partial S}\right)_\theta$$

If the pressure is represented as a function of entropy and volume, $p = p(S, V(S, \theta))$, then the latter expression is $-(\partial p/\partial S)_\theta$, whence it follows (see (8.46)) that

$$\overline{\Delta V\,\Delta S} = \theta\left(\frac{\partial V}{\partial \theta}\right)_p \tag{10.22}$$

This result shows that volume and entropy fluctuations are connected, or correlated. This is understandable, since if the volume of a system increases, the statistical weight of its state, that is, its entropy, also increases.

EXERCISES

1. Write the Poisson distribution (10.2) for large N and $\overline{N}$.

Solution. We represent (10.2) in the form

$$W_N = \frac{1}{(2\pi\overline{N})^{1/2}} \exp\left(-\overline{N} + N \ln \overline{N} - N \ln N + N\right)$$

where $N!$ is written to the same accuracy as in Exercise 1, Section 1. Then we express $\ln(\overline{N}/N)$ as $-\ln[1 + (N - \overline{N})/\overline{N}]$ and expand it in a series up to the second term inclusive. This leads to the Gaussian distribution

$$W_N = \frac{1}{(2\pi\overline{N})^{1/2}} \exp\left[-\frac{1}{2}\frac{(N-\overline{N})^2}{\overline{N}}\right]$$

and the mean quadratic fluctuation $\overline{(\Delta N)^2} = \overline{N}$. The same value of the fluctuation follows from the exact Poisson distribution:

$$\overline{N^2} = \sum N^2 W_N = e^{-\overline{N}}\overline{N}\frac{\partial}{\partial \overline{N}}\left(\overline{N}\frac{\partial}{\partial \overline{N}}\sum \frac{(\overline{N})^N}{N!}\right)$$

$$= e^{-\overline{N}}\overline{N}\frac{\partial}{\partial \overline{N}}\left(\overline{N}\frac{\partial}{\partial \overline{N}}e^{\overline{N}}\right) = (\overline{N})^2 + \overline{N}$$

or

$$\overline{(\Delta N)^2} = \overline{N^2} - (\overline{N})^2 = \overline{N}$$

2. Find the pressure fluctuation at constant entropy and the entropy fluctuation at constant pressure.

Answers.

$$\overline{(\Delta S)^2} = c_p, \qquad \overline{(\Delta p)^2} = -\theta\left(\frac{\partial p}{\partial V}\right)_S$$

3. Find the mean value $\overline{\Delta\theta\,\Delta p}$.

Answer.

$$\overline{\Delta\theta\,\Delta p} = \frac{\theta^2}{c_V}\left(\frac{\partial p}{\partial \theta}\right)_V$$

4. Find the fluctuation of energy and of the number of quanta for an electromagnetic field of a given frequency.

Solution. From the expression

$$\overline{E}_\omega = h\omega\left(e^{h\omega/\theta} - 1\right)^{-1}$$

we obtain with the help of (7.23)

$$\overline{(\Delta E_\omega)^2} = (h\omega)^2 e^{h\omega/\theta} (e^{h\omega/\theta} - 1)^{-2}$$

This formula can be represented as

$$\overline{(\Delta E_\omega)^2} = (h\omega)^2 [(e^{h\omega/\theta} - 1)^{-1} + (e^{h\omega/\theta} - 1)^{-2}]$$

Introducing the number of quanta of a given frequency, $\overline{N}_\omega = E_\omega/h\omega$, we obtain

$$\overline{(\Delta N_\omega)^2} = \overline{N}_\omega + (\overline{N}_\omega)^2$$

The fluctuation of the number of quanta is not like the fluctuation of the number of particles of a Boltzmann gas in Exercise 1.

5. A suspended simple pendulum performs fluctuational oscillations around the equilibrium position. Find the mean square of the angle of deviation from the vertical.

Solution. Denote the length of the pendulum by l and its mass by m. From [Sec. 7] the potential energy of the pendulum in a deflection through an angle φ is equal to $mgl\varphi^2/2$. In this case it is the minimum work appearing in the fluctuation probability. Hence

$$\overline{\varphi^2} = \frac{\theta}{mgl}$$

11

PHASE EQUILIBRIA

Separation into Phases. A substance consisting of molecules of one type is characterized by four quantities: number of particles, temperature, pressure, and volume. Of these, only three quantities are independent, since the equation of state must be satisfied. Thus, for an ideal gas the gas law $pV = N_A\theta$ holds.

An ideal gas uniformly fills the whole of an available volume and in this respect is an exception rather than the rule. For example, if we take one gram of water at 20 °C, no pressure at all can force it to uniformly fill a volume of 10 cm^3 (concerning negative pressures see further on in this section). A gram of water placed in such a volume at 20 °C will separate into two parts, liquid and gaseous; in other words, it does not remain uniform. And a certain, very specific, equilibrium pressure is established in the system.

In a state of statistical equilibrium the mean number of molecules passing from the liquid to the vapour is equal to the mean number of molecules condensing from the vapour to the liquid. It is easy

to see that this condition cannot hold at all pressures: the number of molecules impinging on the liquid surface in unit time is in direct proportion to the pressure, whereas the number of evaporating molecules depends on pressure very weakly. Therefore, at a given temperature pressure alone ensures the equilibrium between liquid and vapour. Under other conditions separation may occur into liquid and solid, into gas and solid, into solids of different crystalline modifications or, in general, into *phases*.

Condition for Phase Equilibrium. Equilibrium pressure as a function of temperature can be expressed in general form by the methods of statistical physics and does not require a detailed examination of the mechanism of the transition from one phase to another.

In equilibrium, the temperature and pressure in both phases is, of course, the same. This condition is necessary, though not sufficient, for equilibrium. In addition to the equality of temperature and pressure in both phases, a necessary condition is that the thermodynamic potential G (see Sec. 8) be minimal. The thermodynamic potential is additive: it equals the sum of the potentials of both phases, and the condition of it being minimal is written as follows:

$$dG = dG_1 + dG_2 = 0 \tag{11.1}$$

At a given temperature and pressure any change in G_1 and G_2 is due solely to changes in the number of particles. From (8.47) we find that

$$dG_1 = \mu_1 \, dN_1, \qquad dG_2 = \mu_2 \, dN_2 \tag{11.2}$$

But since as many molecules leave one phase as enter the other

$$dN_1 = -dN_2$$

we find that

$$(\mu_1 - \mu_2) \, dN_1 = 0 \tag{11.3}$$

Since dN_1 is an arbitrary number, the condition for phase equilibrium reduces to the equality of the chemical potential of both phases:

$$\mu_1 (p, \theta) = \mu_2 (p, \theta) \tag{11.4}$$

This equation can be represented in the form of a curve in the p, θ-plane. In other words, to a certain temperature there corresponds a very definite pressure.

Three phases of one and the same substance may also be in equilibrium. In that case the equilibrium condition is written in the form of two equations:

$$\mu_1 (p, \theta) = \mu_2 (p, \theta) = \mu_3 (p, \theta) \tag{11.5}$$

These equations define a single point in the p, θ-plane (the *triple point*), out of which emerge the equilibrium curves between every two of the three phases (Figure 7).

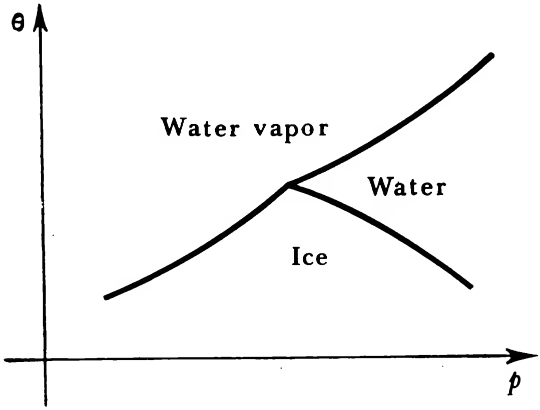


Figure 7

Latent Heat. Usually two phases of the same substance differ greatly: their specific volume, entropy, energy and other additive quantities experience a discontinuity at the transition point. Since the transition occurs at constant pressure, the latent heat is equal to the change in the enthalpy. We shall refer this heat to a single molecule, which means that enthalpy must also be referred to a single molecule (we denote it by h). Similarly, we denote the entropy referred to a single molecule by s. Then the latent heat referred to a single molecule is

$$q = h_2 - h_1 \tag{11.6}$$

In the most general case enthalpy is connected with the thermodynamic potential by the relation $H = G + \theta S$. Going over to quantities referred to a single molecule and applying (8.48), we obtain

$$h = \mu + \theta s \tag{11.7}$$

whence

$$q = \mu_2 - \mu_1 + \theta (s_2 - s_1)$$

But in equilibrium $\mu_1 = \mu_2$, therefore the latent heat is equal to the temperature multiplied by the entropy change in the phase transition:

$$q = \theta (s_2 - s_1) \tag{11.8}$$

This result is quite understandable, since phase transition is a reversible process.

The Clausius-Clapeyron Equation. Consider two phases of the same substance occurring in mutual equilibrium. Suppose the temperature in the equilibrium system has changed somewhat. Let us see how the pressure should change so as to maintain the equilibrium. In other words, we must determine the derivative $dp/d\theta$ along the equilibrium curve.

The dependence of the equilibrium pressure upon temperature is given in the form of an implicit function (11.4). The derivative is found according to the general rule:

$$\frac{dp}{d\theta} = -\left[\frac{\partial(\mu_1 - \mu_2)}{\partial\theta}\right]_p \Big/ \left[\frac{\partial(\mu_1 - \mu_2)}{\partial p}\right]_\theta \tag{11.9}$$

From (8.48)

$$\left(\frac{\partial\mu}{\partial\theta}\right)_p = -s, \qquad \left(\frac{\partial\mu}{\partial p}\right)_\theta = v \tag{11.10}$$

where v is the volume referred to a single molecule. Multiplying the numerator and denominator of the right-hand side of (11.9) by θ and invoking (11.8), we obtain the required equation:

$$\frac{\partial p}{d\theta} = \frac{q}{\theta(v_2 - v_1)} \tag{11.11}$$

which is known as the *Clausius-Clapeyron equation.*

Suppose a transition is considered for which q is positive, for example, melting. The sign of the derivative depends then on the phase that possesses the greater specific volume: liquid or solid. For example, the specific volume of water at the melting point is less than the specific volume of ice, so that $v_2 - v_1$ is a negative quantity. If the pressure above the water-ice system is increased but the equilibrium is not disturbed, the temperature will decrease. Ice, as is known, actually does melt under pressure.

In the transition to the gaseous phase (vapourization in the case of a liquid, sublimation for a solid), we have the inequality $v_2 \gg v_1$. Neglecting v_1 in Eq. (11.11) and replacing v_2 by θ/p, we obtain

$$\frac{d\ln p}{d\ln\theta} = \frac{q}{\theta} \tag{11.12}$$

This derivative is always positive. Therefore the water equilibrium curves close to the triple point can be represented approximately as shown in Figure 7. The equilibrium curve between water and ice has a negative derivative.

A Nearly Ideal Gas. To describe a phase transition with the help of the equation of state we must have an equation that holds for both phases; theory, however, offers no such equation. Use is therefore made of Van der Waals' model equation to construct a quali-

tative picture of the transition in a gas-liquid system. To arrive at this equation it is first useful to investigate a gas only slightly deviating from the ideal and then carry out the necessary extrapolation to a system capable of condensing. Such an approach offers an understanding of the meaning of the parameters defining the degree to which a gas deviates from ideal.

As long as a gas does not differ much from ideal the usual methods of statistics should be applied. Let us begin with the expression for free energy (8.40), assuming that the gas consists of identical particles in classical spatial motion. In that case we must introduce $N!$ into the free energy, as was explained in Section 9. We write it down in the following form:

$$F = -\theta \ln \left(\frac{1}{N!} \int e^{-E'/\theta}\, d\Gamma' \int e^{-U/\theta}\, d\Gamma_0 \right) \qquad (11.13)$$

Here, U is the energy of interaction between individual molecules, which depends on their spatial configuration, E' is the remaining part of the Hamiltonian of the gas, and $d\Gamma_0$ and $d\Gamma'$ are corresponding elements of the phase volume.

The deviation of the gas from ideal is described by the second integral in (11.13). Transform its logarithm as follows:

$$\ln \int e^{-U/\theta}\, d\Gamma_0 = \ln \left[\int (e^{-U/\theta} - 1)\, d\Gamma_0 + \int d\Gamma_0 \right]$$
$$= \ln \left\{ \left[1 + \int (e^{-U/\theta} - 1)\, d\Gamma_0 \Big/ \int d\Gamma_0 \right] \int d\Gamma_0 \right\}$$

Since by definition the gas is nearly ideal, the ratio U/θ in the main part of the phase space should be regarded as a small qauntity, so that the integral of the difference $(e^{-U/\theta} - 1)$ is small in comparison with the integral of 1. In this approximation we obtain

$$\ln \int e^{-U/\theta} d\Gamma_0 = \ln \int d\Gamma_0 + \int (e^{-U/\theta} - 1)\, d\Gamma_0 \Big/ \int d\Gamma_0 \qquad (11.14)$$

If U depends only upon the coordinates of the centres of inertia of the molecules, then $d\Gamma_0 = \prod_{k=1}^{N} dV_k$, so that $\int d\Gamma_0$ is V^N. The value of U may also depend upon the orientation of the molecules in space, but this does not affect the final result. To calculate the second term we again make use of the fact that the gas is nearly ideal, that is, it has low density. In that case the interactions among the molecules occur mainly in collisions of pairs of molecules. Simultaneous encounters of three molecules are highly improbable and make no appreciable contribution to the partition function. The number of all possible pairs of N molecules is $N(N-1)/2 \approx N^2/2$. Integration is possible directly with respect to the coordinates of the centres of

mass of all the other molecules except the colliding pair:

$$\int (e^{-U/\theta}-1)\,d\Gamma_0 \Big/ \int d\Gamma_0$$
$$=\frac{N^2}{2}\times\frac{V^{N-2}}{V^N}\int \left(e^{-U_{12}/\theta}-1\right) dV_1\,dV_2 \qquad (11.15)$$

where U_{12} is the potential energy of interaction of one pair of molecules; U_{12} involves only the relative positions of both molecules, so that integration with respect to the coordinates of their common centre of mass is performed directly. This yields one more factor, V.

The dependence of U_{12} on the distance between the molecules can be described in the following way. Starting with a certain distance r_0 and less, they repulse strongly, like solid spheres. Consequently, at $r \leqslant r_0$ the repulsion energy U is much greater than θ, which means that $e^{-U_{12}/\theta}$ can be neglected in comparison with unity. At $r > r_0$ repulsion turns into attraction. This is an experimental fact, since all gases are capable of condensing, which would have been impossible given only forces of repulsion. In the attraction domain $|U_{12}| \ll \theta$. We arrive at the equation

$$\int (e^{-U_{12}/\theta}-1)\,dV_{12} = -\int\limits_{r\leqslant r_0} dV_{12} - \int\limits_{r>r_0} \frac{U_{12}}{\theta}\,dV_{12}$$
$$\equiv -\beta+\frac{\alpha}{\theta} \qquad (11.16)$$

If we agree to regard r_0 as the double "radius" of the molecule, then β is, apparently, its volume multiplied by a factor of eight.

Substituting the found expression for the integral $\int e^{-U/\theta}\,d\Gamma_0$ into (11.13), we find the free energy of a nearly ideal gas:

$$F = -N\theta \ln\left(\frac{e}{N}V\int e^{-E'/\theta}\,d\Gamma'\right)+\frac{N^2\theta}{2V}\left(\beta-\frac{\alpha}{\theta}\right)$$
$$=F_{\text{ideal}}+\frac{N^2\theta}{2V}\left(\beta-\frac{\alpha}{\theta}\right) \qquad (11.17)$$

Here, the first term is the free energy of an ideal gas, and the second is a correction for its nonideal character proportional to the first power of the gas density N/V. In other words, we have obtained here the first term of the power expansion in the gas density. This could be done because

$$\int U_{12}\,dV_{12} < \infty$$

(the integral converges). For this it is necessary for the attractive forces to decrease faster than the third power of the distance between the molecules. Otherwise the approximation of pair collisions does not hold.

We now find the gas pressure:

$$p=-\left(\frac{\partial F}{\partial V}\right)_\theta=\frac{N\theta}{V}+\frac{N^2\theta}{2V^2}\left(\beta-\frac{\alpha}{\theta}\right) \tag{11.18}$$

This formula holds only for a not very dense gas. Now let us extrapolate to an arbitrary density.

Van der Waals' Equation. At very large compressions, when the mean distance between the centres of molecules is close to r_0, the pressure must become very great, since only a very small free volume remains. This physical requirement can formally be satisfied in the following way. Instead of Eq. (11.18) we write Van der Waals' equation

$$p=\frac{N\theta}{V-N\beta/2}-\frac{N^2\alpha}{2V^2} \tag{11.19}$$

This formula cannot be developed on the basis of statistical mechanics. It is of a purely model nature. However, it takes account of a necessary property of a nonideal gas: the great increase in pressure when the molecules are brought close together. Formally, from Eq. (11.19) the pressure becomes infinite when the gas volume equals four times the volume of all the molecules, $N\beta/2$. The relative nature of the result is also apparent from the fact that the quadruple volume is involved: the factor 4 does not appear explicitly.

Intermolecular attraction reduces the pressure in the gas in proportion to the density of its kinetic energy (see Eq. (2.22)). The kinetic energy of a molecule impringing on a wall is reduced by the attraction of that molecule by the other molecules in the volume. If the attraction is proportional to the number of interacting pairs, then obviously the decrease in kinetic energy per molecule is of the order of N/V, which in terms of the kinetic energy per unit energy yields a quantity of the order $(N/V)^2$, which is taken into account in Van der Waals' equation (11.19).

Conventionally it is written in the following notation:

$$\frac{N\beta}{2}\equiv b, \qquad \frac{N^2\alpha}{2}\equiv a$$

so that finally we obtain

$$p=\frac{N\theta}{V-b}-\frac{a}{V^2} \tag{11.20}$$

The exact equation of state for a real liquid should be much more complex than (11.20), but it would in any case be applicable only to a comparatively narrow class of liquids and could not describe the "liquid-gas" transition in general form.

At the same time, the model equation (11.20) is sufficiently general to describe a phase transition, it is comparatively simple in form,

and in the limiting case of low densities transforms into the ideal gas law. The equation may have to be developed in greater detail only if qualitatively new details of the behaviour of a gas-liquid system not included in Eq. (11.20) are discovered.

The necessary equation of state cannot involve less than two parameters. If there are no forces of attraction ($a = 0$), condensation does not occur, and without the excluded volume b the system will contract into a point instead of into a liquid of finite dimensions.

Van der Waals' Equation and Phase Transition. We shall now show how it follows from Van der Waals' equation that there exists a domain of states in which a substance separates into gas and liquid

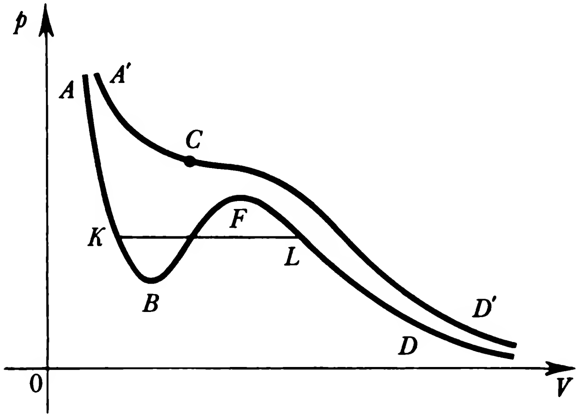


Figure 8

phases. Equation (11.20) is of the third degree with respect to volume. At certain values of θ and p it must have three real roots. The pressure-volume curve at constant temperature (an isotherm) is shown in Figure 8. Between points B and F, the derivative $(\partial p/\partial V)_\theta$ is positive, and from the first inequality in (10.17) the state of a substance with this sign of the derivative is unstable. Hence the substance will necessarily separate into two phases in this region.

The section AK of the curve corresponds to the liquid state, that is, to small volume. As the pressure decreases the liquid expands to point K, after which change occurs along the straight line KL. The points K and L are uniquely defined from the equality for chemical potentials (11.4), and the intermediate points on the straight line correspond to a mixture of the liquid in the state belonging to K and the vapour in the state L. Note that the position of point K at the temperature corresponding to the given isotherm is defined uniquely.

The section KB is not absolutely unstable, since on it $(\partial p/\partial V)_\theta < 0$. The states of this section can be attained without allowing the formation of vapour bubbles in the liquid (a superheated liquid: see Sec. 14). Besides, the liquid must be free from foreign inclusions, for instance, bubbles of other gases. Sometimes the section KL lies partly below the abscissa axis, which corresponds to negative pressure, that is, extension of the liquid. It can indeed extend if it adheres everywhere to the walls of the vessel and has no free surface. The section FL corresponds to a supercooled vapour, which can be obtained if condensation centres are prevented from forming. Such condensation centres, or nuclei, easily arise on ions. This is the principle on which the Wilson chamber for the observation of the tracks of charged particles is based.

The Critical Point. At a sufficiently high temperature the first term on the right in Van der Waals' equation predominates over the second. The equation then becomes very like the ideal gas law for a volume $(V - b)$. But such an equation has only one real root for each value of p. This corresponds to the well-known fact that at high temperatures a substance does not separate into liquid and gaseous phases at any pressure.

Let us find the temperature at which the separation into phases ceases. On the corresponding isotherm $A'CD'$ (Fig. 8), the points B and F, where the derivative $(\partial p/\partial V)_\theta$ becomes zero, merge into one point C, and the domain of unstable states disappears. All three roots of Eq. (11.20) merge at point C, so that C corresponds to a triple root. But if V_C is a triple root, the expansion of the function with respect to the difference $(V - V_C)$ must begin with the third-order term. The linear and quadratic terms turn zero if the first and second pressure derivatives at point C are zero. Hence it is easy to determine the position of point C from Van der Waals' equation.

Write the condition for the first and second derivatives to become zero:

$$\left(\frac{\partial p}{\partial V}\right)_{\theta(V=V_C,\ \theta=\theta_C)} = -\frac{N\theta_C}{(V_C-b)^2}+\frac{2a}{(V_C)^3}=0 \tag{11.21}$$

$$\left(\frac{\partial^2 p}{\partial V^2}\right)_{\theta(V=V_C,\ \theta=\theta_C)} = \frac{2N\theta_C}{(V_C-b)^3}-\frac{6a}{(V_C)^4}=0 \tag{11.22}$$

From this we obtain

$$\frac{V_C-b}{2}=\frac{V_C}{3}$$

or

$$V_C = 3b \tag{11.23}$$

From (11.21) we have

$$a=\frac{N\theta_C\,(V_C)^3}{2\,(V_C-b)^2}=\frac{27}{8}\,N\theta_C b$$

Consequently

$$\theta_C = \frac{8}{27}\frac{8}{Nb} \tag{11.24}$$

From Van der Waals' equation the pressure at point C is

$$p_C = \frac{N\theta_C}{V_C - b} - \frac{a}{(V_C)^2} \tag{11.25}$$

If we represent the phase equilibrium curve in the p, θ-plane, the curve will terminate at the point C ($p = p_C$, $\theta = \theta_C$), which is called the *critical point*. At temperatures $\theta > \theta_C$ separation into phases does not occur.

The critical point can exist only on the equilibrium curve between two phases such that have no features incapable of varying continuously. An example of such a feature is the regularity of a crystal lattice: in principle, the position of an atom in an ideal crystal defines, for a given spatial orientation, the position of the whole crystal, no matter how large it is. But the position of an atom in a liquid affects only the position of its closest neighbours. Therefore, a continuous transition between the solid crystalline phase and liquid phase of some substances is impossible. The curve separating the crystalline and liquid phases has no critical point that could be passed in such a way as to make possible a gradual transformation of the liquid into a crystal without exhibiting a clearly defined melting temperature.

The Law of Corresponding States. Let us eliminate the constants a, b, and N from Eq. (11.20) with the help of Eqs. (11.23)-(11.25):

$$a = 3(V_C)^2 p_C, \qquad b = \frac{V_C}{3}, \qquad N = \frac{8}{3}\frac{p_C V_C}{\theta_C} \tag{11.26}$$

The last of these three expressions shows how much the substance differs from an ideal gas at the critical point: the factor 3/8 appears in the equation of state $p_C V_C = (3/8) N_A \theta_C$ (instead of 1). But as a general rule this equation is not satisfied. Van der Waals' equation is of an approximate nature, and it is therefore not at all surprising that for real substances

$$p_C V_C \neq \frac{3}{8} N_A \theta_C$$

If we now substitute (11.26) into (11.20), we obtain

$$\frac{p}{p_C} = \frac{8\theta/\theta_C}{3(V/V_C) - 1} - 3\left(\frac{V_C}{V}\right)^2 \tag{11.27}$$

Formula (11.27) expresses a special form of the *law of corresponding states*: for two different substances the ratios p/p_C, θ/θ_C, and V/V_C

are related by a single universal equation. It should be noted that in general form the law of corresponding states, especially for substances of similar structure, is satisfied better in practice than the specific formula (11.27) based on Van der Waals' equation. The general law does not impose a definite functional form on the equation of state. But there are, of course, deviations from the law of corresponding states: for two substances having the same ratios p/p_C and θ/θ_C the ratios V/V_C are not strictly the same.

Properties of a Substance Close to the Critical Point. We shall now investigate the properties of a substance close to the critical point in general form, without assuming the validity of Van der Waals' equation (11.20). Let us represent the first factor in the formula for the probability of volume fluctuation as

$$W_V = \exp\left\{\frac{1}{2\theta}(\Delta p\,\Delta V)\right\}$$
$$= \exp\left\{\frac{1}{2\theta}\left[\left(\frac{\partial p}{\partial V}\right)_\theta (\Delta V)^2 + \frac{1}{2}\left(\frac{\partial^2 p}{\partial V^2}\right)_\theta (\Delta V)^3 + \frac{1}{6}\left(\frac{\partial^3 p}{\partial V^3}\right)_\theta (\Delta V)^4\right]\right\}$$

At the critical point $(\partial p/\partial V)_\theta = 0$. But then necessarily $(\partial^2 p/\partial V^2)_\theta = 0$; otherwise at one of the two signs of ΔV the probability of an infinitely great deviation of the volume from the equilibrium value will tend to infinity. The next derivative $(\partial^3 p/\partial V^3)_\theta$ must be negative. In that case the probability of volume fluctuations at the critical point is approximately equal to $\exp\{(1/12\theta)(\partial^3 p/\partial V^3)_\theta(\Delta V)^4\}$ and tends to zero at $\Delta V \to \infty$, which assures stability of the substance at the critical point. That is why the expansion of p on the critical isotherm begins with the term proportional to $(V - V_C)^3$, and the expansion of the derivative begins, accordingly, with the square of the difference, $(V - V_C)^2$.

As for the temperature dependence, the first term of the expansion is linear with respect to $(\theta - \theta_C)$ because the dependence of pressure on temperature does not exhibit any peculiarities close to the critical point.

Thus the expansion with respect to two variables in the critical region has the form

$$\left(\frac{\partial p}{\partial V}\right)_\theta = -\lambda(V - V_C)^2 - \nu(\theta - \theta_C) \tag{11.28}$$

Here $\nu > 0$, because at temperatures above the critical the inequality $(\partial p/\partial V)_\theta < 0$ must always be satisfied: all points in the p,V-plane corresponding to temperatures above θ_C are stable. Accordingly, $\lambda > 0$ too. At temperatures below the critical, $(\partial p/\partial V)_\theta$ vanishes

at two points (B and F in Figure 8). Consequently

$$V_B - V_C = -\left[\frac{\nu}{\lambda}(\theta_C - \theta)\right]^{1/2}$$

$$V_F - V_C = \left[\frac{\nu}{\lambda}(\theta_C - \theta)\right]^{1/2} \qquad (11.29)$$

Let us now find the points K and L on the isotherm which define the phase transition line. For this we make use of the phase equilibrium condition, $\mu_K = \mu_L$. It is conveniently written in the form of an integral taken along the isotherm through points K and L:

$$\int_K^L d\mu = 0 \qquad (11.30)$$

Multiplying by N and then replacing dG at θ = constant by $V\,dp$, we obtain

$$N\int_K^L d\mu = \int_K^L dG = \int_K^L V\,dp = \int (V - V_C)\,dp$$

because from the condition $p_L = p_K$ the integral $\int_K^L dp = p_L - p_K$ vanishes. Now let us use the initial expression (11.28). Then, taking into account that the integration is along the isotherm and replacing dp according to the initial expansion (11.28), we reduce the condition for the equality of the chemical potentials to the following form:

$$\int_{V_K}^{V_L} (V - V_C)\,[\lambda (V - V_C)^2 + \nu(\theta - \theta_C)]\,dV = 0 \qquad (11.31)$$

The integrand is odd with respect to $(V - V_C)$; hence the integral vanishes if the values of $V - V_C$ at integration limits, $(V_K - V_C)$ and $(V_L - V_C)$, are equal in magnitude and opposite in sign. In other words, the volumes V_K and V_L lie at equal distances from V_C but on opposite sides.

We also represent the condition of pressure equality in integral form:

$$\int_K^L dp = \int_{V_K}^{V_L} \left(\frac{\partial p}{\partial V}\right)_\theta dV = 0 \qquad (11.32)$$

Putting (11.28) into this expression and integrating, we obtain

$$\frac{\lambda}{3}(V_L - V_C)^3 + \nu (V_L - V_C)(\theta - \theta_C)$$
$$- \frac{\lambda}{3}(V_K - V_C)^3 - \nu (V_K - V_C)(\theta - \theta_C) = 0$$

Making use of the fact that $(V_K - V_C) = -(V_L - V_C)$, we obtain the required equation

$$\frac{\lambda}{3}(V_L - V_C)^3 + \nu (V_L - V_C)(\theta - \theta_C) = 0$$

whence it follows that

$$V_L - V_C = V_C - V_K = \left[\frac{3\nu(\theta_C - \theta)}{\lambda}\right]^{1/2} \tag{11.33}$$

Thus, close to the critical point, the domain of absolutely unstable states is, in the ratio $1/3^{1/2}$, narrower than the whole domain where phase separation occurs.

Let us now find the latent heat close to the critical point. By definition,

$$Q \approx \theta_C (S_L - S_K) = \theta_C \left(\frac{\partial S}{\partial V}\right)_{\theta_C} (V_L - V_K) \tag{11.34}$$

At the critical point, $(\partial S/\partial V)_{\theta_C}$ retains a finite value equal to $(\partial p/\partial \theta)_{V_C}$. Therefore, the latent heat is proportional to $(\theta_C - \theta)^{1/2}$. At the critical point itself it becomes zero, as could be expected.

The initial expansion (11.28) agrees with Van der Waals' equation, but one cannot be sure that it refers to real gases or liquids. The critical point is a very special point, and in its vicinity an expansion into a Taylor series may not occur. However, when not too close to the critical point, experiments yield the type of dependence of the quantities involved as obtained here.

Phase Transitions of the Second Kind. At the phase transition point the thermodynamic potentials of both phases are equal. The other additive quantities, such as energy, entropy, and volume, experience discontinuities. But there also exist phase transitions in which not the additive quantities themselves are discontinuous but only their derivatives, such as specific heat, compressibility, etc. An example of this kind of transition was given in Section 5, the transition of liquid helium at 2.2 K.

The specific heat at the transition point changes discontinuously. Another example is the transition of iron from the ferromagnetic to the nonferromagnetic state at 770 °C (the *Curie point*).

Phase transitions of the second kind are frequently observed in crystals. In this case they correspond to a certain change in the translational or rotational symmetry of the lattice. Since the form of symmetry cannot change continuously (the property of symmetry either exists or it does not), symmetry always changes discontinuously. If an entropy discontinuity is exhibited at the same time, we have a phase transition of the *first kind*; if the entropy is continuous and the derivatives experience discontinuities, the transition is of the *second kind*.

There are certain relationships between the discontinuities of derivatives on the line of phase transitions of the second kind. They can be established in the following way, proceeding from the continuous character of entropy and volume:

$$\Delta S = S_2 - S_1 = 0, \quad \Delta V = V_2 - V_1 = 0 \tag{11.35}$$

These equations must be differentiated with respect to temperature along the transition line:

$$\Delta\left(\frac{\partial S}{\partial \theta}\right)_p + \Delta\left(\frac{\partial S}{\partial p}\right)_\theta \frac{dp}{d\theta} = 0 \tag{11.36}$$

$$\Delta\left(\frac{\partial V}{\partial \theta}\right)_p + \Delta\left(\frac{\partial V}{\partial p}\right)_\theta \frac{dp}{d\theta} = 0 \tag{11.37}$$

where $dp/d\theta$ denotes the derivative of pressure with respect to temperature along the transition line. Furthermore, from (8.46),

$$\left(\frac{\partial S}{\partial p}\right)_\theta = -\left(\frac{\partial V}{\partial \theta}\right)_p, \quad \text{and} \quad \left(\frac{\partial S}{\partial \theta}\right)_p = \frac{c_p}{\theta}$$

whence, after cancelling out $\Delta\,(\partial S/\partial p)_\theta$,

$$\Delta c_p = \theta\left(\frac{dp}{d\theta}\right)^2 \Delta\left(\frac{\partial V}{\partial p}\right)_\theta \tag{11.38}$$

Thus, along the line of transitions of the second kind the specific heat discontinuity is associated with the compressibility discontinuity. A similar expression is easily derived for the specific heat discontinuity at constant volume.

Sometimes a line of phase transitions of the first kind becomes at a certain point a line of a phase transition of the second kind. If the transition is associated with a change in symmetry, neither line can simply terminate.

EXERCISES

1. Determine the specific heat of one of the phases of a substance along the curve of phase transitions of the first kind.

Solution. From the definition of specific heat,

$$c = \theta\frac{\partial S}{\partial \theta} = \theta\left[\left(\frac{\partial S}{\partial \theta}\right)_p + \frac{dp}{d\theta}\left(\frac{\partial S}{\partial p}\right)_\theta\right]$$
$$= c_p - \frac{q}{\theta(V_2 - V_1)}\left(\frac{\partial V}{\partial \theta}\right)_p$$

2. Show that at the critical point c_p becomes infinite.

Hint. Use the result of Exercise 4 in Section 8 and the condition defining the critical point.

3. Find the discontinuity in specific heat at constant volume along the line of phase transitions of the second kind, expressed in terms of the compressibility discontinuity.

Answer.

$$\Delta c_V = \theta \left(\frac{dV}{d\theta}\right)^2 \Delta \left(\frac{dp}{\partial V}\right)_\theta$$

4. Show that the correction factor for the thermodynamic potential of a nearly ideal gas is equal to the free energy correction factor expressed in terms of p and θ.

Solution. In general form free energy is represented as

$$F = F_{\text{ideal}}(V, \theta) + \delta F(V, \theta)$$

From this we obtain the expression for pressure:

$$p = \left(\frac{\partial F}{\partial V}\right)_\theta = p_{\text{ideal}} + \delta p$$

Then the thermodynamic potential takes the form

$$G = F + pV = F_{\text{ideal}}(V, \theta) + p_{\text{ideal}}\, V - V\left(\frac{\partial\, \delta F}{\partial V}\right)_\theta + \delta F(V, \theta)$$

Here, $F_{\text{ideal}} + p_{\text{ideal}} V$ is the principal term G_{ideal} expressed in terms of the pressure of an ideal gas under the same conditions. Substituting $p + (\partial \delta F/\partial V)_\theta$ for p_{ideal} and taking into account that $(\partial G/\partial p)_\theta = V$, we obtain the required result

$$G = G_{\text{ideal}}(p, \theta) + \delta F(p, \theta)$$

5. Calculate the temperature change in the Joule-Thomson effect in a nearly ideal gas.

Solution. Replacing the constants α and δ with Van der Waals' constants a and b, write the correction factor for the thermodynamic potential:

$$\delta G = p\left(b - \frac{a}{\theta}\right)$$

Here, the result of Exercise 4 has been used. We find δV for the given p:

$$\delta V = \left(\frac{\partial G}{\partial p}\right)_\theta = b - \frac{a}{\theta}$$

The derivative showing the temperature change is

$$\left(\frac{\partial\theta}{\partial p}\right)_H = \frac{1}{c_p}\left(V\frac{\partial\, \delta V}{\partial\theta} - \delta V\right) = \frac{1}{c_p}\left(\frac{2a}{\theta} - b\right)$$

(see Exercise 7, Section 8). At $\theta = 2a/b$ the sign of the derivative changes (the *inversion temperature*). At sufficiently low temperatures it is always positive, so that when pressure decreases, a gas always cools. Inversion is used in liquefying hydrogen, which at room temperature has a negative derivative $(\partial\theta/\partial p)_H$. Prior to expansion in the Joule-Thomson effect hydrogen is cooled below the inversion point.

6. Show that the expansion (11.18) is applicable to a gas consisting of dipole molecules.

Solution. From [Sec. 16], the interaction energy of two dipoles with moments $\mathbf{d}_1$ and $\mathbf{d}_2$ is

$$U_{12}=\frac{r^2(\mathbf{d}_1\cdot\mathbf{d}_2)-3(\mathbf{d}_1\cdot\mathbf{r})(\mathbf{d}_2\cdot\mathbf{r})}{r^5}$$

Although at large distances it decreases as $1/r^3$, the integral with respect to the spatial orientations of the molecules gives zero in the numerator.

12

DILUTE SOLUTIONS

The Thermodynamic Potential of a Dilute Solution. Many of the statistical laws that apply to ideal gases apply to weak solutions and as such have been studied thoroughly. This is because in a weak solution the molecules of the dissolved substance interact among themselves just as little as the molecules of an ideal gas. However, they interact strongly with the surrounding molecules of the solvent, which makes for the differences between a solution and a gas.

Let us now determine the thermodynamic potential of a weak solution. We proceed from the general expression for the free energy in classical statistics:

$$F=-\theta\ln\int e^{-E/\theta}\,d\Gamma \tag{12.1}$$

Here, the unessential factor $(2\pi h)^N$ has been omitted. The integral is taken over all physically different states of the system. Taking into account the identity of all the N molecules of the solvent and all the n molecules of the solute, the classical partition function can be extended over the entire phase space and divided by the total number of permutations of all identical particles. The number of such permutations is $N!n!$.

Now we write the general expression for the thermodynamic potential corresponding to the free energy (12.1):

$$G=F+pV=\left(-\theta\ln\int e^{-E/\theta}\,d\Gamma+\theta\ln N!+pV\right)+\theta\ln n! \tag{12.2}$$

where the integral extends over the whole phase space of the system. Let us expand the expression in parentheses in powers of the small quantity n/N, taking into account that the zero term of the expan-

sion represents the thermodynamic potential G_0 of the pure solvent. Furthermore, we replace $\ln n!$ by $n \ln (n/e)$ according to Stirling's formula. We find that

$$G = G_0 + \frac{n}{N} B(p, \theta, N) + n\theta \ln \frac{n}{e} \tag{12.3}$$

We can refine the dependence of $B(p, \theta, N)$ upon the number of solvent particles by noting that the thermodynamic potential is an additive function of N and n. In other words, if N and n increase by a certain factor, G must also increase by the same factor. But in (12.3) this requirement is directly satisfied only by the potential of the pure solvent, G_0, which is equal to $N\mu_0$ with μ_0 the chemical potential of the pure solvent. For the second and third terms also to be additive, first write the third term in the form

$$n\theta \ln \frac{n}{e} = n\theta \ln \frac{n}{eN} + n\theta \ln N$$

After this the thermodynamic potential takes the form

$$G = N\mu_0(p, \theta) + n\theta \ln \frac{n}{eN} + n\left[\frac{B(p, \theta, N)}{N} + \theta \ln N\right]$$

In order to obtain an additive expression we must require that the function $(B/N + \theta \ln N)$ be totally independent of N. The result is a general expression for the thermodynamic potential of a dilute solution:

$$G = N\mu_0(p, \theta) + n\theta \ln \frac{n}{eN} + nX(p, \theta) \tag{12.4}$$

The chemical potential of the solvent is

$$\mu = \frac{\partial G}{\partial N} = \mu_0 - \frac{n\theta}{N} \tag{12.5}$$

and the chemical potential of the solute is

$$\mu' = \frac{\partial G}{\partial n} = \theta \ln \frac{n}{N} + X(p, \theta) \tag{12.6}$$

Osmotic Pressure. There exist semipermeable membranes through which the molecules of a solvent pass freely but the molecules of the dissolved substance (solute) cannot pass. The solvent on both sides of such the membrane must be in equilibrium when in unit time equal numbers of molecules pass through the membrane both ways. The equilibrium condition is that the chemical potential of the pure solvent on one side of the membrane is equal to the chemical potential of the solvent in the solution on the other. The temperature on both sides is the same, for otherwise statistical equilibrium simply cannot be established. Only the pressure can differ,

provided the pressure difference is restrained by the membrane. Denoting the pressure difference by Δp, we obtain the equilibrium condition

$$\mu_0(p, \theta) = \mu(p+\Delta p, \theta) = \mu_0(p+\Delta p, \theta) - \frac{n\theta}{N} \tag{12.7}$$

Let us expand μ_0 in a power series in Δp, restricting ourselves to the linear term (such an expansion is justified since for a liquid Δp is a small quantity):

$$\mu_0(p+\Delta p, \theta) = \mu_0(p, \theta) + \frac{\partial\mu_0}{\partial p}\Delta p \tag{12.8}$$

But $(\partial\mu_0/\partial p)$ is equal to the volume of pure solvent per one molecule:

$$\frac{\partial\mu_0}{\partial p} = \frac{V}{N}$$

whence we obtain the equation

$$V\Delta p = n\theta \tag{12.9}$$

The excess pressure Δp in the solution is called the *osmotic pressure*. Equation (12.9) is in every way similar to the ideal gas law. It was originally found experimentally and was the basis for formulating the thermodynamic theory of solutions. Here we have developed it from the general principles of statistics.

Phase Equilibrium of a Solvent (Raoult's Laws). We shall now consider another case, when equilibrium is also established over the molecules of the solvent. Imagine a solution in equilibrium with another phase of the solvent, into which the solute does not pass. Let us determine the displacement of the phase equilibrium curve in the p,θ-plane.

We denote the chemical potential of the phase into which the solute does not pass as μ_1. Then the phase equilibrium condition for the pure solvent is given by the equation

$$\mu_1(p, \theta) = \mu_0(p, \theta) \tag{12.10}$$

The equilibrium of the other phase of the solvent with the solution is displaced and is given by the following condition:

$$\mu_1(p+\Delta p, \theta+\Delta\theta) = \mu_0(p+\Delta p, \theta+\Delta\theta) - \frac{n\theta}{N} \tag{12.11}$$

Let us expand the chemical potentials in a series in Δp and $\Delta\theta$:

$$\begin{aligned} &\mu_1(p+\Delta p, \theta+\Delta\theta) - \mu_0(p+\Delta p, \theta+\Delta\theta) \\ &= \mu_1(p, \theta) - \mu_0(p, \theta) + \frac{\partial(\mu_1-\mu_0)}{\partial p}\Delta p + \frac{\partial(\mu_1-\mu_0)}{\partial\theta}\Delta\theta \\ &= (v_1 - v_0)\Delta p - (s_1 - s_0)\Delta\theta \end{aligned} \tag{12.12}$$

Assume now that the pressure in the system is the same as above the pure solvent, that is, $\Delta p = 0$. Then the displacement of the equilibrium temperature, $\Delta\theta$, is

$$\Delta\theta = \frac{n\theta^2}{Nq} = \frac{n\theta^2}{Q} \tag{12.13}$$

where $Q = N\theta(s_1 - s_0)$ is the latent heat of phase transition of the pure solvent. For vapourization $Q > 0$, hence if the solute does not pass into vapour, then $\Delta\theta > 0$, that is, the equilibrium temperature rises. Indeed, a solution has a higher boiling point than the pure solvent.

Now assume that the solute does not pass into the solid phase of the solvent. In that case Q is the heat of solidification, $Q < 0$. It is seen from this that the melting point of a solution is lower than that of the pure solvent. The use of cooling mixtures is based on this property of solutions.

Now consider equilibrium at a given temperature ($\Delta\theta = 0$). In this case the lowering of pressure is determined from (12.12):

$$\Delta p = -\frac{n\theta}{(v_1 - v_0)N} \tag{12.14}$$

If a solution is in equilibrium with its vapour, then $v_1 > v_0$. The product Nv_1 is the volume of the entire solvent in the vapour state. If the solute could be transformed into vapour together with the solvent, the partial pressure of the molecules of the substance would be equal to the lowering of the equilibrium pressure over the solution. The relative lowering of the pressure, $-\Delta p/p$, is equal to the concentration of the solution, n/N.

Solute Equilibrium. A solution is said to be *saturated* if it is in equilibrium with the solute. The equilibrium condition consists in that the chemical potential of the pure solute μ_0' is equal to its chemical potential in the dissolved state:

$$\mu_0' = \mu' = \theta \ln \frac{n_0}{N} + X(p, \theta) \tag{12.15}$$

It is assumed here that the saturated solution is also dilute, that is, $n_0 \ll N$.

If the pure substance is in a gaseous state its chemical potential depends upon the pressure according to the law (9.17):

$$\mu_0' = \theta \ln p + f_1(\theta) \tag{12.16}$$

The dependence of the function $X(p, \theta)$ upon the external pressure is relatively weak: $X(p, \theta)$ is determined by the properties of the condensed phase, which do not change when the external pressure varies within a few atmospheres. Equating the right-hand sides

of (12.15) and (12.16) and taking antilogarithms, we find that the equilibrium concentration of the dissolved gas is proportional to its pressure above the liquid (*Henry's law*):

$$\frac{n_0}{N} = a(\theta)\, p \tag{12.17}$$

The factor of p weakly depends upon pressure.

Heat of Dilution. The *heat of dilution* at constant pressure is the difference between the enthalpies of the substances comprising the solution before and after being dissolved (Sec. 8). The enthalpy is related to the thermodynamic potential in the following way:

$$H = G - \theta\left(\frac{\partial G}{\partial \theta}\right)_p = -\theta^2\left(\frac{\partial}{\partial \theta}\frac{G}{\theta}\right)_p \tag{12.18}$$

Therefore the heat of dilution is

$$Q = -\theta^2 \frac{\partial}{\partial \theta}\frac{1}{\theta}\left(N\mu_0 + n\theta \ln \frac{n}{eN} + nX - n\mu_0' - N\mu_0\right) \tag{12.19}$$

where μ_0' is the chemical potential of the solute. The quantities appearing in the equation can be expressed in terms of the concentration of the saturated solution n_0/N with the help of the saturation condition (12.15). This yields

$$Q = -n\theta^2 \frac{\partial}{\partial \theta} \ln \frac{n}{n_0} \tag{12.20}$$

The heat of dilution per one molecule is $Q/n = q$, or

$$q = -\theta^2 \frac{\partial}{\partial \theta} \ln \frac{n}{n_0} = \frac{\theta^2}{n_0}\frac{\partial n_0}{\partial \theta} \tag{12.21}$$

Thus, if the concentration of a saturated solution increases with temperature, then heat is absorbed in the dissolution process.

The Le Chatelier-Braun Principle. Suppose that heat is supplied to a saturated solution in equilibrium with the solute. Then, if $(\partial n_0/\partial \theta) > 0$, part of the substance continues to dissolve, and heat is expended not only in raising the temperature but in dissolving as well. But if $(\partial n_0/\partial \theta) < 0$, part of the substance leaves the solution, on which, according to (12.21), heat is also expended. In both cases changes take place in an equilibrium system which counteract the external action contributing to the rise of the temperature. This example illustrates a general rule, known in thermodynamics as the *Le Chatelier-Braun principle*.

The Phase Rule. Suppose we have k substances (components) distributed in the form of solutions of arbitrary concentration in f

phases. How many parameters define the equilibrium state of such a system?

The chemical potentials of substances depend upon temperature, pressure, and the relative concentrations. The concentrations of all the components in any phase are related by the identities

$$\sum_{l=1}^{k} c_l^f = 1 \tag{12.22}$$

since by definition

$$c_l^f = n_l^f \Big/ \sum_{l'=1}^{k} n_{l'}^f$$

The equilibrium condition consists in the equality of the chemical potentials of each of the k substances over all phases:

$$\begin{aligned} \mu_1^1 (p, \theta, c_1^1, c_2^1, \ldots, c_k^1) &= \mu_1^2 (p, 0, c_1^2, c_2^2, \ldots, c_k^2) \\ &= \ldots = \mu_1^f (p, \theta, c_1^f, c_2^f, \ldots, c_k^f) \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \mu_k^1 (p, \theta, c_1^1, c_2^1, \ldots, c_k^1) &= \mu_k^2 (p, \theta, c_1^2, c_2^2, \ldots, c_k^2) \\ &= \ldots = \mu_k^f (p, \theta, c_1^f, c_2^f, \ldots, c_k^f) \end{aligned} \tag{12.23}$$

Here, the superscript always denotes the phase, and the subscript the substance.

Equation (12.23) involves k concentrations in f phases plus two more variables, temperature and pressure, giving a total of $kf + 2$ variables.

For each substance there are $(f - 1)$ Eqs. (12.23) and an additional f equations (12.22), or in all $k(f - 1) + f$ equations for determining $kf + 2$ variables. The number of independent variables which can vary arbitrarily is equal to the difference between the number of variables and the number of equations, that is,

$$r = kf + 2 - k(f - 1) - f = k - f + 2 \tag{12.24}$$

The quantity r is called the number of thermodynamic degrees of freedom of a system. Equation (12.24) expresses the *Gibbs phase rule*: the number of degrees of freedom equals the number of components minus the number of phases plus two. For example, if two phases of the same substance are in equilibrium, then $r = 1$; in such a system one variable, temperature or pressure, can be changed arbitrarily. In a two-component and two-phase system there are two degrees of freedom: the concentration of the components in one of the phases can be varied together with the temperature or pressure.

EXERCISES

1. Determine the change in the volume of substances in the formation of a dilute solution.

Answer.

$$\Delta V = -n\theta \frac{\partial \ln n_0}{\partial p}$$

This formula is in agreement with the general Le Chatelier-Braun principle.

2. Write the thermodynamic potential of a dilute solution of two substances in one solvent in an approximation that makes it possible to take account of the effect of one substance on the solubility of the other.

Solution. Carry out an expansion in a series up to the terms quadratic with respect to concentration:

$$G = N\mu_0(p, \theta) + n_1\theta \ln \frac{n_1}{eN} + n_2\theta \ln \frac{n_2}{eN} + n_1X_1 + n_2X_2$$

$$+ \frac{n_1^2}{2N} Y_{11} + \frac{n_1n_2}{N} Y_{12} + \frac{n_2^2}{2N} Y_{22}$$

From this follow the expressions for the chemical potentials of both dissolved components:

$$\mu_1' = \theta \ln c_1 + X_1 + c_1Y_{11} + c_2Y_{12}$$
$$\mu_2' = \theta \ln c_2 + X_2 + c_{12}Y_{12} + c_{22}Y_2$$

Taking this into account and using an equation analogous to (12.15), we obtain the final result.

13

CHEMICAL EQUILIBRIA

Irreversible and Reversible Reactions. Like all processes whose rates do not coincide with the rates of change of the external parameters of the systems involved, chemical reactions that take place with a finite rate are irreversible. For example, the burning of oxyhydrogen gas irreversibly produces water vapour.

If a certain quantity of the oxyhydrogen mixture is prepared in a closed vessel, the state of the mixture will be thermodynamically unstable with respect to the reaction. True, the reaction by no means proceeds directly according to the "gross equation" $2H_2 + O_2 =$

$= 2H_2O$. For that the molecules would have to overcome very high potential barriers (see Exercise, 2, Section 2). Actually, the reaction must proceed through stages involving intermediate unstable substances OH, H, O with unsaturated valencies, the *active centres*.

The initial formation of active centres is very difficult, and an oxyhydrogen mixture can be kept at room temperature indefinitely. But if active centres have been somehow produced, for instance, by a powerful electric spark or contact with an open flame, they renew and multiply in the course of the reaction (a chain reaction)[10]. In these conditions, when the active centres multiply at a sufficiently high rate, the reaction proceeds explosively.

But chemical reactions never go on to the end. If the explosion is produced in a sufficiently strong vessel (bomb), the final equilibrium state will include hydrogen, oxygen, and water vapour in concentrations depending on the initial concentrations of the mixture, as well as on the temperature and pressure. This final state is called chemical equilibrium.

When a state changes slowly, the equilibrium shifts to one side or the other, that is, the quantities of initial or end products may increase. But these chemical reactions around the equilibrium point proceed at the same rate as the rate of change of the external parameters. Consequently, such reactions are reversible, like all processes whose rate is not established spontaneously and is always equal to the rate of change of the quantities determining the equilibrium state of the system.

Condition of Chemical Equilibrium. The state of thermodynamic equilibrium in general and chemical equilibrium in particular can be found with the help of the thermodynamic functions of the substances taking part in the reaction. In this, only the "gross equation" is required, quite irrespective of what intermediate substances the reaction yields, a fact which helped formulate the theory of chemical equilibria back in the nineteenth century. At the same time, investigation of the rates and mechanisms of chemical reactions is being developed to this day. In many reactions a branching, or multiplication, of active centres occurs. However, owing to experimental difficulties, the mechanism of such reactions has not yet been established.

At given temperature and pressure chemical equilibrium is attained only when the thermodynamic potential in the reacting mixture has a minimum, that is,

$$dG = 0 \tag{13.1}$$

---

[10] Most chain reactions are associated with active centres. This was established by N. N. Semenov and his pupils.

At p = constant and θ = constant, the minimum condition for G has the form

$$dG = \sum_i \mu_i \, dN_i = 0 \tag{13.2}$$

Here, μ_i is the chemical potential of the ith substance appearing in the reaction equation. In the oxyhydrogen mixture, for example, these substances are hydrogen, oxygen, and water vapour, all three being in the molecular, nondissociated state. The quantities dN_i are not arbitrary: they change in the course of the reaction and are therefore related by the reaction equation. In other words, N_i can vary only in equivalent (stoichiometric) quantities. For example, in the reaction

$$2CO + O_2 = 2CO_2$$

$dN_{CO} \div dN_{O_2} \div dN_{CO_2} = -2 \div -1 \div 2$. In the reaction of thermal dissociation of hydrogen

$$H_2 = 2H$$

$dN_{H_2} \div dN_H = -1 \div 2$. In general, the number dN_i is proportional to the equivalent of the given substance, ν_i. Equation (13.2) can be rewritten as follows:

$$\sum_i \mu_i \nu_i = 0 \tag{13.3}$$

This equation expresses the condition for the chemical equilibrium of a system.

The Law of Mass Action. Equation (13.3) is especially useful when the explicit expression of the chemical potential of the reacting substances is known, as, for example, in a weak solution or an ideal gas. In the latter case, the equilibrium concentrations of the substances can be determined if we have sufficient information concerning the structure of all the molecules in equilibrium.

The chemical potential of a certain gas in a mixture of ideal gases is, according to (9.17),

$$\mu_i = -\theta \ln \frac{\theta f_i(\theta)}{p_i} \tag{13.4}$$

where $f_i(\theta)$ is the partition function taken over all momentum values of the molecule as a whole, as well as over all its rotational, vibrational, electron and nuclear-spin states. The electron states are essential when they lie close to the ground state of the molecule and far from the dissociation limit. If they lie closer to the dissociation limit, the molecule decomposes before such highly excited states can in any way affect the partition functions (see Exercise 2).

Substituting the expression for chemical potential into the chemical equilibrium condition (13.3) and cancelling out θ, we obtain

$$\sum_i \nu_i \ln p_i = \sum_i \nu_i \ln \theta f_i$$

Taking antilogarithms, we obtain the equilibrium condition, expressed in terms of the partial pressures, from the formula

$$\prod_i p_i^{\nu_i} = \prod_i (\theta f_i)^{\nu_i} \equiv K \qquad (13.5)$$

This equation can also be written in terms of the relative concentrations of the substances by replacing the partial pressures with the help of (9.15):

$$\prod_i c_i^{\nu_i} = p^{-\sum_i \nu_i} \prod_i (\theta f_i)^{\nu_i} = p^{-\sum_i \nu_i} K \qquad (13.6)$$

Here, c_i denotes the concentration of the ith component of the mixture:

$$c_i = \frac{N_i}{N} \qquad (13.7)$$

The pressure in the right-hand side of (13.6) has still to be expressed in terms of the initial pressure or the initial density, which can easily be done from the ideal gas law, taking into account the change in the number of particles as compared with their initial number for the given equilibrium intensity of the chemical reaction.

The equilibrium concentrations of the components depend on the initial quantities of the initial substances involved in the reaction. Thus, the equilibrium concentrations also depend on these quantities, or masses. That is why Eq. (13.6) is also called the *law of mass action.*

The quantity appearing on the right-hand side of Eq. (13.5) is called the *equilibrium constant* of the given reaction, because it does not involve the concentrations of the mixture.

Heat of Reaction. The heat of a chemical reaction taking place at constant pressure is defined as the difference between the enthalpies of the reacting substances before and after the reaction. This heat is conveniently written in terms of a single elementary act of the reaction (see (12.18)):

$$q = \delta_I = -\theta^2 \frac{\partial(\delta G \theta)}{\partial \theta} \qquad (13.8)$$

But in an elementary act $\delta G = \sum_i \mu_i \nu_i$, and the heat of reaction is

$$q = -\theta^2 \frac{\partial}{\partial \theta}\left(\sum_i \mu_i \nu_i / \theta\right) \qquad (13.9)$$

This expression denotes the heat *absorbed* in a reaction. The heat *liberated* would have to be denoted with the opposite sign.

When the law of mass action applies, the heat of reaction is expressed in terms of the equilibrium constant K:

$$q = \theta^2 \frac{\partial \ln K}{\partial \theta} \tag{13.10}$$

Equation (13.10) agrees with the Le Chatelier-Braun principle, which can be easily observed from the following reasoning. If $(\partial \ln K/\partial\theta) > 0$, then as the temperature increases the equilibrium tends towards a predominance of those substances that are involved in the reaction equation with positive coefficients ν_i. The concentrations of these substances appear in the numerator of the left-hand side of Eq. (13.6). But then, from (13.10), the system absorbs heat and the reactions in it oppose any increase in temperature. By increasing or reducing the temperature of an equilibrium system we may cause reversible reactions to proceed in any direction we wish.

EXERCISES

1. Write the equations of the law of mass action for the reaction $2CO + O_2 = 2CO_2$ if initially a moles of CO and b moles of O_2 were involved.

Solution. Suppose x moles of O_2 took part in the reaction. In that case $2x$ moles of CO were involved, and $2x$ moles of CO_2 were formed. In all there are $a + b - 3x + 2x = a + b - x$ moles of different gases in the system. Their concentrations are

$$c_{CO} = \frac{a - 2x}{a + b - x}$$

$$c_{O_2} = \frac{b - x}{a + b - x}$$

$$c_{CO_2} = \frac{2x}{a + b - x}$$

The equilibrium equation appears thus:

$$\frac{(2x)^2 (a + b - x)}{(a - 2x)^2 (b - x)} = K(p, \theta)$$

where p is the equilibrium pressure, which differs from the pressure of the original substances at the same temperature by the factor $(a + b - x)/(a+b)$. From this we obtain the equation for the required quantity x:

$$\frac{x^2}{(a - 2x)^2 (b - x)} = \frac{p_0 K}{4(a + b)}$$

2. Calculate the equilibrium constant for the thermal dissociation of nitrogen using the following data.

The ground state of a nitrogen atom is ^4S. The first excited state, ^2D, lies 2.4 eV higher; the next, ^2P, lies 3.5 eV above the ground state.

The formation energy of an N_2 molecule, referred to absolute zero, is 9.76 eV. The moment of inertia of the ground state of the molecule $I = 13.84 \times 10^{-40}$ g-cm². The vibrational quantum of the molecule is 0.287 eV. In the ground state the orbital angular momentum and the spin angular momentum of the electrons do not have projections on the line joining the nuclei. The first excited state lies more than 6 eV above the ground state of the molecule.

Solution. The partition function for the atoms is

$$f_N = \frac{(2\pi m_N \theta)^{3/2}}{(2\pi h)^3} (4 + 2 \times 5e^{-2.4/\theta} + 2 \times 3e^{-3.5/\theta})$$

Here and further on θ is conveniently expressed in electron-volts, taking into account that 1 eV = 11.600 K. We shall confine ourselves to temperatures for which the partition function for a molecule contains only the ground electron state. Then we obtain (see (9.21))

$$f_{N_2} = \frac{(2\pi m_{N_2}\theta)^{3/2}}{(2\pi h)^3} \frac{2\pi I\theta}{(2\pi h)^2} \frac{4\pi e^{9.76/\theta}}{2[1-\exp(-h\omega/\theta)]}$$

From (13.5) the equilibrium constant for the reaction $N_2 = 2N$ is

$$K = \frac{\theta f_N^2}{f_{N_2}} = (4 + 10e^{-2.4/\theta} + 6e^{-3.5/\theta})^2 \frac{1}{8Ih} \left(\frac{m_N\theta}{\pi}\right)^{3/2}$$
$$\times [1-\exp(-h\omega/\theta)]\, e^{-9.76/\theta}$$

To illustrate, let us find the fraction of dissociated molecules if the temperature is equal to 1 eV and there are 2.7×10^{19} molecules per cubic centimetre. The equation of the law of mass action then appears as

$$\frac{4x^2}{1-x} = 0.494 \times 10^6 \times 5.90 \times 10^{-5} = 27.4$$

Here, the factor before the exponential function is equal to 5×10^6, while the exponential function itself is 5.9×10^{-5}. The equilibrium dissociation is $x = 0.88$. Thus, 88% of all the molecules have already dissociated when the temperature is only one-tenth of the dissociation energy. The relative predominance of the pre-exponential factor over the exponential function at such comparatively low temperatures is due to the fact that the statistical weight of the dissociated state is determined by the entire volume occupied by the gas, while the nondissociated state is determined only by the volume of the molecules, which is why dissociation is already highly probable at atmospheric gas density (2.7×10^{19} mol-cm^{-3}).

It is sometimes said figuratively that here "entropy works against energy": what is nonadvantageous from the point of view of energy becomes highly probable thanks to the increase in entropy, that is, in statistical weight. In the temperature range in which most of the dissociation occurs

the specific heat of the gas increases considerably, because most of the applied heat is expended on the dissociation of molecules.

3. Find the degree of thermal ionization of helium as a function of temperature and pressure. Disregard second ionization.

Solution. The ionization potential of helium is 24.47 eV, while the first excited state lies 20.5 eV above the ground state.

Ionization equilibrium satisfies the law of mass action:

$$\frac{c_e c_{\mathrm{He}^+}}{c_{\mathrm{He}}}=\frac{K}{p}$$

Here the partition functions are

$$f_e=2\frac{(2\pi m_e\theta)^{3/2}}{(2\pi h)^3}$$

$$f_{\mathrm{He}^+}=2\frac{(2\pi m_{\mathrm{He}}\theta)^{3/2}}{(2\pi h)^3}$$

$$f_{\mathrm{He}}=\frac{(2\pi m_{\mathrm{He}}\theta)^{3/2}}{(2\pi h)^3}e^{24.47/\theta}$$

(the factors 2 take account of the spins of the He⁺ ion and the electron).

From this we can express the equilibrium constant as follows:

$$K=4\frac{(2\pi m_e\theta)^{3/2}}{(2\pi h)^3}\theta\times e^{-24.47/\theta}=\frac{1}{h^3}\left(\frac{2m_e^3\theta^5}{\pi^3}\right)^{1/2}e^{-24.47/\theta}$$

If the initial pressure of helium is p_0, the equation of ionization equilibrium takes the form

$$\frac{x^2}{1-x}=\frac{K}{p_0}$$

Here, as in the previous problem, the pre-exponential factor predominates over the exponential function, which is equal to 2.19×10^{-3}, owing to the large statistical weight of the ionized state. The excited states of the helium atom make a very small contribution to the partition function. At higher temperatures the first ionization is complete, so that there are simply no neutral atoms capable of being excited.

14

SURFACE PHENOMENA

The Thermodynamic Potential of a Surface. So far we have considered only bulk properties of matter, and all our findings with respect to phase and chemical equilibria in solutions refer, strictly speaking, to very large systems.

Surfaces separating different substances or different phases of the same substance exhibit special properties, which depend on both the nature and the states of the adjoining bodies.

The thermodynamic potential per unit surface of contact of two media depends on the temperature θ and the pressure p in the surrounding media. In equilibrium, θ and p are constant over the whole surface. Interaction between the contacting sections of the phase separation boundary occurs across these boundaries. Therefore, different sections can be treated as quasi-independent subsystems: the areas of individual sections are proportional to the square of their dimensions, while the lengths of the separation lines are proportional to the first power of their dimensions. If the sections are large enough, interaction within the surface is stronger than interaction along a line. Hence, the thermodynamic potential of a surface is additive for the same reason that a volume potential is. If the potential per unit surface of two media is denoted α, and the surface area ζ, by virtue of additivity the potential of the whole surface is

$$G = \alpha\zeta \tag{14.1}$$

Surface Tension. The work done at constant pressure and temperature is equal to the change in thermodynamic potential (Sec. 7). Therefore, the work done in a unit change of surface area is equal to α. This work is called the *surface tension* of two given media.

It is easy to demonstrate the relation between this definition of surface tension and its elementary definition. Let a film of liquid be stretched on a rigid rectangular wire frame with one movable side. If the length of the movable member is unity, then a force acts on it from the side of the film, equal to twice the surface tension of the film (because the film has two sides). In a displacement of the movable member over unit length the force of surface tension does work numerically equal to double its magnitude. But in this the film surface increases by two units, so that the work done in increasing the surface by unity is in fact equal to the "force" of surface tension.

When the surface area increases, part of the atoms pass from the body of the liquid to the surface layer; for this they must overcome some of the pull exerted by other atoms. This explains the origin of the work that is lost (or gained, depending on the nature of the contacting volumes) in increasing the surface area. The surface tension of a condensed phase at a boundary with a vacuum is, of course, always positive.

In equilibrium, the thermodynamic potential has a minimum. In this case the minimum is attained simply at the least surface area ζ. Therefore, a liquid film stretched on a nonplanar frame assumes the least possible surface area for the given frame. A liquid drop in ideal equilibrium assumes a spherical shape, which has the least surface area for the given volume.

Heat of Surface Increase. When surface area increases, not only work is done but heat is transferred as well. Since the process of increasing the surface area is reversible, the heat is determined from the general formula (8.18) $\theta = \theta\Delta S$ in terms of surface entropy (see (8.46)). Substituting the thermodynamic potential (14.1) into this formula, we obtain the expression for the heat of surface increase:

$$\theta = -\theta(\zeta_2 - \zeta_1)\frac{\partial\alpha}{\partial\theta} \tag{14.2}$$

Heat may be gained or lost, depending upon the sign of $(\partial\alpha/\partial\theta)$.

The Equilibrium of Vapour Above a Drop. The phase equilibrium condition changes if we take into account the surface thermodynamic potential as well as the volume potential. Of course, the general condition $dG = 0$ holds, but it no longer reduces to the form (11.4), $\mu_1 = \mu_2$. It can be written down in the following general form:

$$\frac{\partial G_1}{\partial N} = \frac{\partial G_2}{\partial N} \tag{14.3}$$

Let the subscript 1 refer to a vapour phase contained in a large volume, and the subscript 2 to a small liquid drop of radius R. Then, for the first phase,

$$\frac{\partial G}{\partial N} = \mu_1 \tag{14.4}$$

and for the second phase

$$\frac{\partial G}{\partial N_2} = \mu_2 + \alpha\frac{\partial\zeta}{\partial N} \tag{14.5}$$

The derivative in the second term is calculated in the following way:

$$\alpha\frac{\partial\zeta}{\partial N} = 8\pi\alpha R\frac{\partial R}{\partial N} \tag{14.6}$$

If the density of the liquid is ρ mol-cm^{-3}, then $R = N^{1/3}(4\pi\rho/3)^{-1/3}$ and

$$\frac{\partial R}{\partial N} = \frac{1}{3}\frac{R}{N} \tag{14.7}$$

Substituting this into (14.6) and expressing N in terms of R, we obtain

$$\alpha\frac{\partial\zeta}{\partial N} = \frac{8\pi\alpha R^2}{3(4/3)\pi R^3\rho} = \frac{2\alpha}{\rho R} \tag{14.8}$$

Thus, the equilibrium condition between the vapour and a liquid drop is expressed by the equation

$$\mu_1(p, \theta) = \mu_2(p, \theta) + \frac{2\alpha}{\rho R} \tag{14.9}$$

Let us represent the pressure p as $p_0 + \Delta p$ (where Δp is the equilibrium pressure above a plane surface). The expansion of the chemical potential in powers of Δp yields (see (12.12))

$$(v_1 - v_2)\,\Delta p = \frac{2\alpha}{\rho R} \tag{14.10}$$

Neglecting the specific volume of the liquid compared to the specific volume of the vapour, we find the final expression for the excess pressure:

$$\Delta p = \frac{2\alpha}{R\rho v_1} = \frac{2\alpha v_2}{R v_1} = \frac{2\alpha v_2}{R\theta}\, p \tag{14.11}$$

The formula for the pressure inside a vapour bubble is obtained similarly, but with opposite sign.

Stability of Supersaturated Phases. Thus, the equilibrium pressure of vapour above the convex surface of a drop is greater, and above the concave surface of a bubble less, than above a plane surface. This explains the relative stability of supersaturated phases mentioned in Section 11. If a liquid drop appears spontaneously in a supersaturated vapour, and its radius R is

$$R < \frac{2\alpha p}{\rho\theta\,(p' - p)} \tag{14.12}$$

(where p is the equilibrium vapour pressure over a plane surface, and p' is the pressure of the supersaturated vapour), the drop evaporates again. Further condensation on it is highly improbable, since this is a fluctuation phenomenon. Only if the inequality (14.12) is reversed can the drop begin to grow. But the spontaneous formation of a large drop, like any major fluctuation, is highly improbable. That is why condensation usually begins on small nuclei already present in the vapour, for example, on ions.

In exactly the same way we can explain why a highly purified superheated liquid does not boil. Boiling of a liquid consists in the formation of vapour bubbles within the liquid. For a bubble to keep from collapsing under the external pressure of the liquid the equilibrium vapour pressure must be at least equal to the external atmospheric pressure above the liquid. But if the vapour equilibrium pressure above a plane surface is only equal to atmospheric pressure, the pressure inside the bubble is not sufficient for equilibrium. Therefore, a bubble that is too small cannot grow.

APPENDIX TO PART I

An integral of the form

$$\int_0^\infty x^n (e^x \pm 1)^{-1}\, dx$$

is calculated in the following way. The function $(e^x \pm 1)^{-1}$ is expanded in a series of powers of e^{-x}:

$$(e^x \pm 1)^{-1} = \sum_{k=1}^{\infty} (\pm 1)^{k+1} e^{-kx}$$

This series is integrated term by term, the integral for each individual term being represented thus:

$$\int_0^\infty e^{-kx} x^n\, dx = \frac{1}{k^{n+1}} \int_0^\infty e^{-z} z^n\, dz$$

If n is an integer, this integral is equal to $n!$, as can be easily shown by integrating by parts. At n half-integral it can be evaluated according to the formulas derived in Exercise 3 of Section 1. For instance, substituting $z^{1/2} = u$, we obtain

$$\int_0^\infty e^{-z} z^{1/2}\, dz = 2 \int_0^\infty e^{-u^2} u^2\, du = \frac{\pi^{1/2}}{2}$$

and in general

$$\int_0^\infty e^{-z} z^{m-1/2}\, dz = 2 \int_0^\infty e^{-u^2} u^{2m}\, du = \frac{1 \times 3 \times 5\ \ldots\ (2m-1)\pi^{1/2}}{2^m}$$

We shall call this quantity $(m - 1/2)!$, so that in general

$$\int_0^\infty e^{-z} z^n\, dz = n!$$

Hence

$$\int_0^\infty z^n (e^z \pm 1)^{-1}\, dz = n! \sum_{k=1}^{\infty} (\pm 1)^{k+1} \frac{1}{k^{n+1}}$$

The summation with respect to the upper sign (plus) can be reduced to the summation with respect to the lower sign (minus):

$$1 - \frac{1}{2^{n+1}} + \frac{1}{3^{n+1}} - \frac{1}{4^{n+1}} + \ldots$$

$$= 1 + \frac{1}{2^{n+1}} + \frac{1}{3^{n+1}} \ldots + \frac{(-2)}{2^{n+1}} \left(1 + \frac{1}{2^{n+1}} + \frac{1}{3^{n+1}} + \ldots\right)$$

$$= \left(1 - \frac{1}{2^n}\right) \left(1 + \frac{1}{2^{n+1}} + \frac{1}{3^{n+1}} + \frac{1}{4^{n+1}} + \ldots\right)$$

Finally, the summation involving positive signs has the following values:

n	1/2	1	3/2	2	5/2	3
$\sum_{k=1}^{\infty} \frac{1}{k^{n+1}}$	2.612	1.645	1.341	1.202	1.127	1.082

For odd n we have the following formulas:

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}, \qquad \sum_{k=1}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{90}$$

Therefore

$$\int_0^{\infty} \frac{x^3\,dx}{e^x - 1} = 3! \sum_{k=1}^{\infty} \frac{1}{k^4} = \frac{\pi^4}{15}$$

We also note that

$$\int_0^{\infty} \frac{x^{1/2}\,dx}{e^x - 1} = \frac{\pi^{1/2}}{2} \sum_{k=1}^{\infty} \frac{1}{k^{3/2}} = \frac{\pi^{1/2}}{2} \times 2.612 = 2.31$$

PART II

HYDRODYNAMICS AND GAS DYNAMICS

15

THE GENERAL EQUATIONS OF HYDRODYNAMICS

The mechanics of a continuous medium is, by its very essence, a statistical department of theoretical physics, insofar as it investigates the motions of large assemblies of atoms and molecules. However, in most applications we need not take into account the atomic structure of matter and can treat it as a continuous medium. In the statistical sense this corresponds to a transition to mean values, that is, the substitution of statistically averaged quantities for real, fluctuating ones.

The Stress Tensor. One such quantity, pressure, that is, the mean linear momentum carried in unit time across unit surface, was examined in Sections 2 and 8. Pressure is exerted not only on the walls of the vessel containing a substance. It is the same in any cross section of the volume. In liquids and gases at rest it acts perpendicular to any surface drawn through the fluid or on its boundary. But in a moving fluid tangential components of momentum may also be carried across a surface, generating tangential forces.

Obviously, like any vector lying on a surface, such forces have two components. Thus, the force which a given volume exerts on another volume adjoining it across the surface separating them is described in the most general case by three quantities: a normal component and two tangential components.

Imagine a volume in the shape of a parallelepiped with sides dx, dy, and dz cut out of a continuous medium. Face $dS_x = dy\,dz$ is perpendicular to the x axis, etc. Acting on unit area of this face are three components of the force, or stresses, as they are conventionally

called: p_{xx}, p_{xy}, and p_{xz}. The first subscript of these quantities is the same as for dS, in this case x, and the second subscript states the direction of the force. The other components p_{yx}, p_{yy}, p_{yz}, p_{zx}, p_{zy}, and p_{zz} are defined similarly.

It is easy to show that the nine quantities p_{ik} form a tensor of rank 2. For this we must calculate the resultant force in the x direction applied to a volume of arbitrary form. Let $d\mathbf{S}$ be an element of a surface containing a volume. Then acting on the face $d\mathbf{S}$ in the x direction is a force $dS_x p_{xx} + dS_y p_{yx} + dS_z p_{zx}$, where dS_x, dS_y, and dS_z are the projections of $d\mathbf{S}$. Indeed, a flow of any quantity p_{ix} across face $d\mathbf{S}$ with components dS_i is expressed in the same way as the flow of a vector A_i across that face. The additional subscript x is of no consequence. Hence, the resultant force df_x is, according to the summation rule given in [Sec. 2],

$$df_x = dS_x p_{xx} + dS_y p_{yx} + dS_z p_{zx} = dS_i p_{ix}$$

$$f_x = \int dS_i p_{ix} \tag{15.1}$$

and in general

$$df_k = dS_i p_{ik}$$

Since df_k and dS_i are components of a vector, the p_{ik}'s connecting them represent the components of a tensor of rank 2[1]. Its diagonal components are analogous to pressure, but defined with the opposite sign. Thus, for conventional pressure subject to Pascal's law one simply has to write $p_{ik} = -p\delta_{ik}$. Then the force acting on face $d\mathbf{S}$ is defined as $df_k = p\, dS_i \delta_{ik} = -p\, dS_k$. This force is normal to the face.

We shall show that the tensor p_{ik} is symmetric. For that we must calculate the moment of the forces acting on a cubical element. Let one of its apexes coincide with the origin of the coordinate system and the three adjoining sides be coincident with the coordinate axes (Figure 9). Let us determine the projection of the moment of the forces with respect to the x axis. Only the stress components applied to faces $ABCD$ and $EFCD$ have arms equal respectively to dy and dz. The force in the x direction applied to face $ABCD$ is $p_{yz}\, dS_y = p_{yz}\, dx\, dz$. Its moment about the x axis is equal to $p_{yz}\, dx\, dz\, dy$. The moment of the force applied to face $EFCD$ is equal to $-p_{zy} \times dx\, dy\, dz$. The minus is there because, as can be seen in the figure, this moment causes a rotation in the opposite direction of the rotation caused by the force applied to face dS_y. The resultant moment of force is thus equal to $(p_{yz} - p_{zy})\, dx\, dy\, dz$.

---

[1] See [9.6]. The set of nine coefficients linking the components of two vectors form a tensor of rank 2.

The other components p_{ik} do not produce a resulting moment of force. The moment of force thus found is equal to the product of the angular acceleration, $\ddot{\varphi}$, of the cube and the moment of inertia, I (assuming that the fluid body contains no other carriers of angular momentum, such as tiny tops). But we assume from the outset that the medium is homogeneous and contains no macroscopic inclusions. Then the moment of inertia I is two orders smaller than the volume of the small element since it additionally includes the square of the

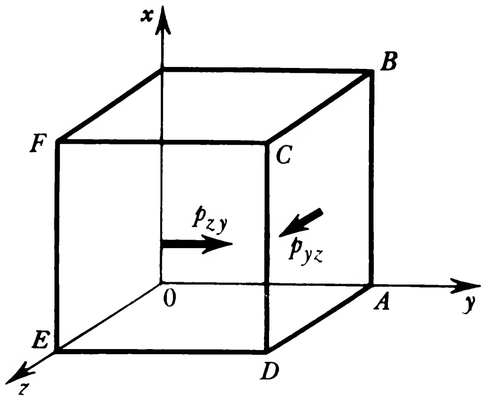


Figure 9

arm. Therefore if we equate the moment of force to the product of the moment of inertia and the angular acceleration and divide both sides of the equation by the volume element, on the left-hand side there is left a zero-order quantity and on the right-hand side the square of the arm mentioned before, that is, a small quantity of the second order. This is possible only if $p_{yz} - p_{zy} = 0$ or, in general form, if

$$p_{ik} = p_{ki} \tag{15.2}$$

Any tensor of rank 2, p_{ik}, can be identically represented as a sum of three components:

$$p_{ik} = \frac{1}{3}\delta_{ik}p_{ll} + \left(\frac{p_{ik}+p_{ki}}{2} - \frac{1}{3}\delta_{ik}p_{ll}\right) + \frac{p_{ik}-p_{ki}}{2} \tag{15.3}$$

In going over to a new coordinate system each of these components undergoes a transformation in terms of the respective components in the old system: the scalar p_{ll} in terms of a scalar, the symmetric component in terms of a symmetric component, and the antisymmetric component in terms of an antisymmetric component. The scalar

is analogous to pressure in Pascal's law, but with an opposite sign. The second component appears in media that resist changes in form at constant volume. The third component, as was shown, is simply zero.

The General Equations of Motion of a Continuous Medium. An element of mass contained in some volume dV is $\rho\, dV$, where ρ is the density of the mass. If the velocity of this element is $\mathbf{v}$, then according to Newton's Second Law

$$\frac{d\mathbf{v}}{dt}\rho\, dV = d\mathbf{F}$$

Integrated over a finite volume, this equation has the form

$$\int \frac{d\mathbf{v}}{dt}\rho\, dV = \int d\mathbf{F}$$

The force comprises two components. Firstly, there may be a given body force of density $\rho\mathbf{f}$. Secondly, acting on the volume element are neighbouring elements; this is described with the help of the stress tensor p_{ik}. Their joint action on the whole body is determined from (15.1). Therefore, in general form Newton's Second Law for an arbitrary volume element is written as follows:

$$\int \frac{dv_i}{dt}\rho\, dV = \int \rho f_i\, dV + \int p_{ik}\, ds_k \tag{15.4}$$

The integral over the surface is transformed into an integral over the volume according to Gauss' theorem:

$$\int p_{ki}\, dS_k = \int \frac{\partial p_{ki}}{\partial x_k} dV$$

Since Eq. (15.4) holds for an arbitrary volume, the integrands are equal:

$$\rho \frac{dv_i}{dt} = \frac{\partial p_{ki}}{\partial x_k} + \rho f_i \tag{15.5}$$

The unknown quantities here are density, the three velocity components and six stress components, a total of ten quantities for three equations. In the most general case, consequently, a solution of the problem requires seven more equations specifying the properties of the fluid.

Mass is always conserved (at nonrelativistic velocities). The corresponding equation is written exactly like the charge conservation law [12.18]:

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \rho\mathbf{v} = 0 \tag{15.6}$$

Let us consider one of the simplest continuous media.

Ideal Liquid or Gas. In fluid mechanics, a liquid or gas is called *ideal* if all the nondiagonal components of the stress tensor are zero. This property is invariant with respect to rotations of the coordinate system only if all the diagonal components are equal, that is, when the stress tensor becomes a scalar. Since a scalar does not transform in rotations of a coordinate system, pressure is perpendicular to any surface element whatever its spatial orientation. This is always true of a liquid at rest. An ideal liquid is one in which this property (Pascal's law) is conserved in motion as well:

$$p_{ik} = -p\delta_{ik} \tag{15.7}$$

where p, as mentioned before, is the pressure within the liquid. This reduces the number of unknown quantities to five, so that one more equation is needed in addition to (15.5) and (15.6) (the body force f_k is treated as an external, that is, given, force).

To obtain the required equation we must determine the type of thermodynamic process the fluid flow conforms to. If the flow is not too slow, individual elements of the moving fluid have no time to exchange heat. It is transmitted comparatively slowly through molecular motion, and the basic energy exchange between various body elements occurs through work done in compression or expansion. In this book we shall consider only flow in which there is no heat exchange.

Heat transfer from hot to cold points is but one of several possible irreversible processes. In fluid motion there may also be internal friction, or chemical reactions may be taking place. If these irreversible processes do not occur, the flow can be considered isentropic.

When the initial entropy of a liquid is constant throughout the whole volume, the equation of an isentropic process defines the pressure as a function of density, or the density as a function of pressure:

$$p = p(\rho) \qquad \text{or} \qquad \rho = \rho(p) \tag{15.8}$$

In the more general case we must write the condition for the constancy of the entropy of a body element:

$$\frac{d}{dt}(S\rho\, dV) = 0$$

Here S is the entropy referred to unit mass. Since $\rho\, dV$ (the body element) is, for a liquid, a constant quantity, we obtain

$$\frac{dS(\mathbf{r}, t)}{dt} = \frac{\partial S}{\partial t} + \frac{d\mathbf{r}}{dt}\operatorname{grad} S = \frac{\partial S}{\partial t} + \mathbf{v}\operatorname{grad} S \tag{15.9}$$

where $d\mathbf{r}/dt = \mathbf{v}$ is the velocity of the given volume element.

The total derivative of $\mathbf{v}$ is computed similarly to (15.9) (see [11.31]):

$$\frac{d\mathbf{v}}{dt} = \frac{\partial \mathbf{v}}{\partial t} + \frac{dx}{dt}\frac{\partial \mathbf{v}}{\partial x} + \frac{dy}{dt}\frac{\partial \mathbf{v}}{\partial y} + \frac{dz}{dt}\frac{\partial \mathbf{v}}{\partial z} = \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v}\cdot\nabla)\,\mathbf{v} \tag{15.10}$$

Substituting (15.7) and (15.10) into (15.5), we obtain the equation of motion of an ideal liquid (the *Euler equation of motion*):

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t}+(\mathbf{v}\cdot\nabla)\,\mathbf{v}\right)=-\operatorname{grad} p+\rho\mathbf{f} \tag{15.11}$$

Together with (15.6), (15.8) or (15.9) this gives a complete set of equations of flow for an ideal compressible (ρ is not constant!) fluid.

The Law of Conservation of Energy. Let us now see what form the conservation laws take in such motion. For simplicity we shall consider that the body force has a potential U, that is, $\mathbf{f} = -\operatorname{grad} U$, with U depending on the coordinates only and not on time. We divide (15.11) by ρ and multiply the result scalarly by $\mathbf{v}$. Then on the left-hand side we have the total derivative

$$\mathbf{v}\,\frac{d\mathbf{v}}{dt}=\frac{d}{dt}\,\frac{v^2}{2}$$

The expression on the right-hand side of (15.11) also involves a total derivative, $\mathbf{v} \operatorname{grad} U = dU/dt$, since U is not an explicit function of time. The term $\mathbf{v} \operatorname{grad} p$ is represented as $(dp/dt - \partial p/\partial t)$. But dp/ρ is, from the thermodynamic relationship (8.31), the differential of the enthalpy, dH. Indeed, the total derivative of enthalpy is $dH = \theta\, dS + V\, dp$.

Since the motion is isentropic ($dS = 0$) and V is the volume referred to unit mass ($V = 1/\rho$), $dH = dp/\rho$. Hence

$$\frac{d}{dt}\left(\frac{v^2}{2}+H+U\right)=\frac{1}{\rho}\,\frac{\partial p}{\partial t}$$

We expand the total derivative in the left-hand side and again multiply ρ into the equation:

$$\rho\frac{\partial}{\partial t}\left(\frac{v^2}{2}+H+U\right)+\rho\mathbf{v}\operatorname{grad}\left(\frac{v^2}{2}+H+U\right)=\frac{\partial p}{\partial t} \tag{15.12}$$

Referring now to (15.6), which is called the *continuity equation*, and multiplying it by $v^2/2 + H + U$, we obtain

$$\left(\frac{v^2}{2}+H+U\right)\frac{\partial \rho}{\partial t}+\left(\frac{v^2}{2}+H+U\right)\operatorname{div}\rho\mathbf{v}=0$$

Adding this to (15.12),

$$\frac{\partial}{\partial t}\rho\left(\frac{v^2}{2}+H+U\right)+\operatorname{div}\rho\mathbf{v}\left(\frac{v^2}{2}+H+U\right)=\frac{\partial p}{\partial t}$$

and taking advantage of the fact that by (8.13) $H = E + pV = E + p/\rho$, where E is the energy of unit mass of the substance, we obtain $\rho H = \rho E + p$.

Substituting this into the last equation, we finally obtain an equation analogous in form to the mass conservation equation, or the

continuity equation (15.6):

$$\frac{\partial}{\partial t}\rho\left(\frac{v^2}{2}+E+U\right)+\operatorname{div}\rho\mathbf{v}\left(\frac{v^2}{2}+H+U\right)=0 \tag{15.13}$$

Thus, the energy density consists of three components: the density of the kinetic energy $\rho v^2/2$, the density of the mean, or internal (thermodynamic), energy ρE, and the density of the potential energy ρU. As for the density of the energy flux, it involves not E but the enthalpy H. This means that energy is not only transferred by fluid flow but is also transferred from one volume element to another through the work done in compression. As can be seen from (15.13), the mechanical quantity $\rho v^2/2 + U$ is involved together with the thermodynamic energy E in the total energy balance.

The Law of Conservation of Linear Momentum. In investigating this law the specific properties of the medium and the motion (in the thermodynamic sense) are immaterial. This means that we can proceed from Eq. (15.5), where we must, of course, assume the external force $\mathbf{f}$ equal to zero. Otherwise the momentum could not be conserved.

Rewrite the continuity equation (15.6) in tensor form and multiply by v_i:

$$v_i\frac{\partial\rho}{\partial t}+v_i\frac{\partial\rho v_k}{\partial x_k}=0$$

Also represent the left-hand side of (15.5) in tensor form:

$$\rho\frac{dv_i}{dt}=\rho\left(\frac{\partial v_i}{\partial t}+v_k\frac{\partial}{\partial x_k}v_i\right)$$

Now add the two equations to get

$$\frac{\partial}{\partial t}\rho v_i+\frac{\partial}{\partial x_k}(\rho v_i v_k-p_{ik})=0 \tag{15.14}$$

This equation expresses the law of conservation of linear momentum. The momentum density is ρv_i, and the density of the momentum flux is $(\rho v_i v_k - p_{ik})$. In the flow of an ideal liquid the stress tensor is replaced by pressure according to (15.7), and Eq. (15.14) takes the form

$$\frac{\partial}{\partial t}\rho v_i+\frac{\partial}{\partial x_k}(\rho v_i v_k+\delta_{ik}p)=0 \tag{15.15}$$

It is important that momentum is carried through space not only by the fluid motion, to which the term $\rho v_i v_k$ corresponds, but also by the stress forces or pressure.

Bernoulli's Equation (Weak Form). There is one more, specifically hydrodynamic, conservation law. In the steady isentropic flow of an ideal liquid in a conservative force field the equations of

motion take the form

$$\frac{d\mathbf{v}}{dt} = (\mathbf{v}\cdot\nabla)\,\mathbf{v} = -\frac{\operatorname{grad} p}{\rho} - \operatorname{grad} U \tag{15.16}$$

The steady-state character of the flow is taken into account by putting the partial derivative $\partial\mathbf{v}/\partial t$ equal to zero.

Multiplying (15.16) scalarly by $\mathbf{v}$, we obtain

$$\frac{d}{dt}\left(\frac{v^2}{2} + H + U\right) = 0$$

The total derivative indicates that the differentiated quantity is associated with a given volume element of the liquid (a partial derivative is taken only for a volume element related to a fixed coordinate system).

The fact that the total derivative is equal to zero indicates that the quantity under the differentiation sign is conserved in a volume element in steady isentropic flow:

$$\frac{v^2}{2} + H + U = \text{constant} \tag{15.17}$$

This is *Bernoulli's theorem* in its so-called *weak form.*

The Conservation of Velocity Circulation. Let us draw a closed line through the particles of a liquid or a gas. Conditions in the fluid are the same as were just assumed in the derivation of Bernoulli's theorem, with the exception of that of steady flow.

Denote an element of length along the closed line by $d\mathbf{l}$. The theorem to be proved states that the *velocity circulation* along a closed path in a fluid,

$$\Gamma \equiv \oint \mathbf{v}\, d\mathbf{l} \tag{15.18a}$$

is constant provided the motion is isentropic and the forces are conservative.

Let a denote a coordinate taken along the closed path, thus defining a particle of the fluid. Instead of $d\mathbf{l}$ we write $(d\mathbf{l}/da)\, da$. Then the circulation can be written as

$$\Gamma \equiv \oint \mathbf{v}\, \frac{d\mathbf{l}}{da}\, da \tag{15.18b}$$

The total derivative with respect to time is

$$\frac{d\Gamma}{dt} = \oint \frac{d\mathbf{v}}{dt}\frac{d\mathbf{l}}{da}\, da + \oint \mathbf{v}\,\frac{d}{dt}\frac{d\mathbf{l}}{da}\, da \tag{15.19}$$

We substitute $-(\operatorname{grad} p)/\rho - \operatorname{grad} U$ for $d\mathbf{v}/dt$. The scalar product of this quantity and $d\mathbf{l}$ involves the change in p and U from particle to particle, that is, $-dp/\rho - dU$, because the path is not stationary

in space and moves together with the fluid. The ratio dp/ρ can be replaced by dH, as in (15.12). Differentiations with respect to t and a are interchangeable since they are carried out with respect to independent variables. The given value of a belongs all the time to the same particle. Therefore

$$\frac{d}{dt}\frac{d\mathbf{l}}{da}=\frac{d}{da}\frac{d\mathbf{l}}{dt}=\frac{d\mathbf{v}}{da}, \qquad \mathbf{v}\frac{d\mathbf{v}}{da}=\frac{d}{da}\frac{v^2}{2} \tag{15.20}$$

As a consequence the right-hand side of Eq. (15.19) is reduced to the form

$$\frac{d\Gamma}{dt}=\oint\frac{d}{da}\left(\frac{v^2}{2}-H-U\right)da \tag{15.21}$$

But after traversing the path the a coordinate returns to its initial value and so do the values of v, H, and U. From this it is apparent that

$$\frac{d\Gamma}{dt}=0, \qquad \Gamma=\text{constant} \tag{15.22}$$

If circulation Γ is not zero, then there exists a closed path tangential at all points to the direction of the velocity vector (like the closed line of magnetic induction). Circulation in a fluid takes place along such a closed path. The smoke rings puffed out by expert smokers are the visible lines of the vector curl $\mathbf{v}$. The lines are closed because div curl $\mathbf{v} \equiv 0$. The circulation of the velocity takes place along lines meshed with the ring (like the induction lines of a magnetic field mesh with direct-current lines).

The conditions for the application of the circulation theorem break down in a nonconservative force field. For example, in the nonconservative Coriolis force field [8.7] of the earth's atmosphere there is the circulation of air masses (the trade winds). Nonisentropic flow also causes circulation. Due to the heating of masses of water or air by external sources, pressure ceases to be a unique function of density. In the most general case of a single-phase, one-component medium any thermodynamic quantity is a function of two others. But then dp/ρ is not a total differential and is not expressed in terms of dH.

Thus, if mechanical equilibrium is disturbed or made unstable by unequal heating, circulation occurs in a fluid.

When the conditions for the application of theorem (15.22) are satisfied, circulation cannot appear spontaneously (if there was none before). But then

$$\text{curl } \mathbf{v} = 0$$

which is satisfied at all points of the fluid because the circulation of $\mathbf{v}$ along a closed path transforms according to Stokes' theorem (see [Sec. 11]), into a flow of the curl of $\mathbf{v}$ across the surface stretched over

the closed line:

$$\Gamma = \oint \mathbf{v}\, d\mathbf{l} = \int \operatorname{curl} \mathbf{v}\, d\mathbf{s}$$

Irrotational flow remains irrotational. That means we can introduce the velocity potential

$$\mathbf{v} = \operatorname{grad} \varphi \tag{15.23}$$

where φ is a unique function of coordinates and time; the flow is termed *potential flow*. In the case of rotational flow the vortex lines pass through the same particles of the fluid and are, so to say, attached to them: according to (15.22) the circulation persists along a closed path.

Potential flow possesses a special integral of the motion. From the general formula [11.32] we have

$$\frac{1}{2} \operatorname{grad} v^2 = \mathbf{v} \times \operatorname{curl} \mathbf{v} + (\mathbf{v} \cdot \nabla)\, \mathbf{v}$$

In irrotational flow only the second term remains on the right-hand side. Assuming further that p is a function only of ρ, we reduce the general equation (15.11) to the form

$$\operatorname{grad} \left(\frac{\partial \varphi}{\partial t} + \frac{v^2}{2} + H + U \right) = 0$$

or

$$\frac{\partial \varphi}{\partial t} + \frac{v^2}{2} + H + U = \text{constant} \tag{15.24}$$

This is true over the whole volume of the liquid (the *Lagrange-Cauchy theorem*).

In steady flow $\partial\varphi/\partial t = 0$, and we obtain

$$\frac{v^2}{2} + H + U = \text{constant} \tag{15.25}$$

over the whole volume.

This is the so-called *strong form* of *Bernoulli's theorem*. The quantity which in the weak form was conserved only along the streamline is in this case constant over the whole volume and is thus independent of the coordinates. It is completely defined by its value at one point.

The density of a fluid can often be assumed to be constant to a very good approximation, that is, the fluid is regarded as *incompressible*. The relevant conditions will be formulated more precisely in the next section. Here, we shall only write down the corresponding hydrodynamic equations:

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\, \mathbf{v} = -\frac{1}{\rho} \operatorname{grad} p + \mathbf{f} \tag{15.26}$$

$$\operatorname{div} \mathbf{v} = 0 \tag{15.27}$$

The system (15.26)-(15.27) is complete since it involves four equations and four unknown quantities, v and p. There is no difficulty in writing the integral form of these equations. The department of hydrodynamics that treats of highly compressible fluids is called *gas dynamics* (Secs. 20-25). Weak, or acoustic, compressions are investigated in acoustics.

Equation (15.27) is used to solve problems on steady flow in an incompressible fluid provided the appropriate boundary conditions are stated. Thus, on a fixed solid wall the normal velocity component must be zero. Introducing the velocity potential (15.23), from (15.27) we obtain the equation

$$\nabla^2\varphi = 0 \tag{15.28}$$

for the boundary condition $v_n = \text{grad}_n\, \varphi = 0$ (where grad_n is the normal component of the gradient). On a free surface the pressure must be assumed constant or zero, which is the same thing since only grad p enters the equations. The free surface passes through the streamlines. These conditions are in many cases sufficient, provided we also know the boundary conditions at infinity.

Complex Potential. We shall now consider the case of steady plane flow of an ideal, incompressible fluid. All the streamlines are parallel to a certain plane, which we denote x, y, so that the potential φ depends only upon x and y. From (15.23)

$$v_x = \frac{\partial\varphi}{\partial x}, \qquad v_y = \frac{\partial\varphi}{\partial y} \tag{15.29}$$

In this case the problem of determining the velocity field is greatly simplified by applying functions of a complex variable.

Take a complex function $w \equiv \psi + i\varphi$ depending upon a complex variable $z \equiv x + iy$:

$$w = w(z)$$

The variables x and y are independent. Hence, in the most general case the value of the derivative dw/dz may depend on the differentials dx and dy involved in $dz = dx + i\,dy$, that is, upon the direction of vector dz in the complex plane. The function $w(z)$ is termed *analytical* if the derivative dw/dz does not depend on that direction[2]. Let us determine the conditions that must in this case be imposed upon ψ and φ.

---

[2] It can be shown that in that case the function w can be expanded in a Taylor series in the vicinity of point z.

Write the differentials dw at constant x and y:

$$(dw)_x = \left(\frac{\partial\psi}{\partial x} + i\,\frac{\partial\varphi}{\partial x}\right) dx \tag{15.30a}$$

$$(dw)_y = \left(\frac{\partial\psi}{\partial y} + i\,\frac{\partial\varphi}{\partial y}\right) dy \tag{15.30b}$$

For there to be a limit dw/dz independent of x and y (separately) the factors of dx and $i\,dy$, as well as of $i\,dx$ and dy, in the differentials (15.30a, b) must be equal:

$$\frac{\partial\psi}{\partial x} = \frac{\partial\varphi}{\partial y}, \qquad \frac{\partial\varphi}{\partial x} = -\frac{\partial\psi}{\partial y} \tag{15.31}$$

(the *Cauchy-Riemann equations*). If these are satisfied,

$$\begin{aligned} dw &= \left(\frac{\partial\psi}{\partial x} + i\,\frac{\partial\varphi}{\partial x}\right)(dx + i\,dy) \\ &= \left(\frac{\partial\psi}{\partial x} + i\,\frac{\partial\varphi}{\partial x}\right) dz \equiv \left(\frac{\partial\varphi}{\partial y} - i\,\frac{\partial\psi}{\partial y}\right) dz \end{aligned}$$

that is, a unique limit dw/dz exists.

Eliminating ψ from (15.31), we obtain

$$\frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} = 0 \tag{15.32}$$

so that the function φ can be taken as the plane flow potential. The same holds for ψ.

From the Cauchy-Riemann equations we also obtain the following relation:

$$\frac{\partial\varphi}{\partial x}\,\frac{\partial\psi}{\partial x} + \frac{\partial\varphi}{\partial y}\,\frac{\partial\psi}{\partial y} = 0 \tag{15.33}$$

This means that the gradients of φ and ψ are mutually perpendicular. But in that case the lines of constant values of φ and ψ are also mutually perpendicular, so that grad φ is directed along the line $\psi =$ = constant and grad ψ along $\varphi =$ constant. Consequently, on a solid wall $\psi =$ constant since in that case the vector grad $\varphi = \mathbf{v}$ has no component normal to the wall.

The rectilinear grid of mutually perpendicular lines $x =$ constant and $y =$ constant is mapped into a curvilinear grid $\varphi =$ constant, $\psi =$ constant, and these curves are also mutually perpendicular. For this reason the mapping $w = w(z)$ is called *conformal*, that is, it retains the form of infinitesimal sections of the mapped planes.

Note that φ and ψ can be interchanged: the lines $\psi =$ constant can be taken as equipotential lines and $\varphi =$ constant as the streamlines, which corresponds to changed boundary conditions.

The flow of a viscous fluid (Sec. 17) around a solid body may differ considerably from the potential flow described here. But in superfluid liquid helium potential flow holds rigidly (see Sec. 19). In addition, in some sections of the flow of a real fluid the picture closely approximates potential flow.

EXERCISES

1. A sphere of radius a is immersed in an ideal, incompressible fluid in steady, parallel flow with a velocity $\mathbf{v}_0$. Determine the pattern of potential flow around the obstacle.

Solution. The potential satisfies the Laplace equation $\nabla^2\varphi = 0$. For an undisturbed flow it is equal to $\varphi_0 = \mathbf{v}_0\mathbf{r}$. We shall find the potential of the disturbance caused by the obstacle in the form $\varphi_1 = A\ (\mathbf{v}_0\cdot\nabla)\ r^{-1}$. Since the operators ∇ and ∇^2 are permutable and $\mathbf{v}_0$ is a constant vector, φ_1 satisfies the Laplace equation, $\nabla^2\varphi_1 = 0$, as does φ_0. The velocity is equal to grad φ:

$$\mathbf{v} = \mathrm{grad}\,(\varphi_0+\varphi_1) = \mathbf{v}_0 - A\,\frac{3\,(\mathbf{v}_0\cdot\mathbf{r})\,\mathbf{r}-\mathbf{v}_0 r^2}{r^5}$$

The constant A must be chosen so as to satisfy the boundary condition $v_n = \mathrm{grad}_n\,\varphi = 0$ on the surface of the sphere:

$$v_n = \frac{\mathbf{v}_0\mathbf{r}}{a} - A\,\frac{3\,(\mathbf{v}_0\cdot\mathbf{r})-(\mathbf{v}_0\cdot\mathbf{r})}{a^4} = 0$$

whence $A = a^3/2$. Finally

$$\mathbf{v} = \mathbf{v}_0 - \frac{a^3}{2}\,\frac{3\,(\mathbf{v}_0\cdot\mathbf{r})\,\mathbf{r}-\mathbf{v}_0 r^2}{r^5}$$

The disturbance of the flow created by the sphere at $r > a$ has the form of a field produced by a dipole [16.19].

2. Investigate flows described by the complex potential

$$\psi + i\varphi = w = \frac{\Gamma}{2\pi i}\ln z = \frac{\Gamma}{2\pi i}\ln\,(x+iy)$$

assuming both φ and ψ as the velocity potential.

Solution. We go over to polar coordinates in the x,y-plane:

$$z = re^{i\alpha}$$

Taking the logarithm and separating the imaginary part, we obtain

$$\varphi = \frac{\Gamma}{2\pi}\ln r$$

The velocity has only a radial component

$$v_r = \frac{\partial\varphi}{\partial r} = \frac{\Gamma}{2\pi r}$$

The discharge through a circle centred at the origin of the coordinate system is

$$\Gamma=\int_0^{2\pi} v_r r\, d\alpha$$

The discharge is the same across any closed curve C encompassing the origin (Figure 10) because div $\mathbf{v}=0$ at any point but the origin. It can be seen that the discharge across a closed line not encompassing the origin is zero.

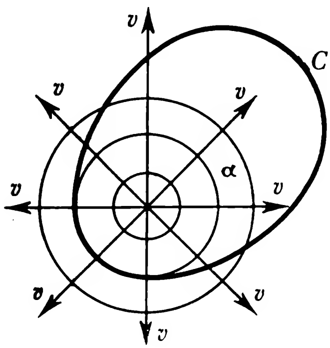


Figure 10

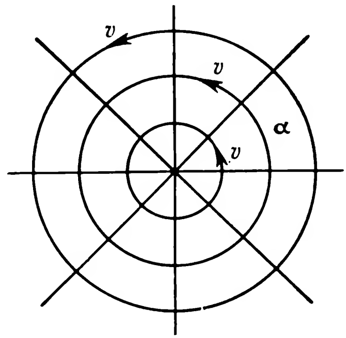


Figure 11

The flow pattern corresponds to a filament source with a constant discharge Γ per unit length perpendicular to the x,y-plane.

But if ψ is taken for the velocity potential, the equipotential lines lie along the radii (Figure 11), since

$$\psi=\frac{\Gamma\alpha}{2\pi}$$

It follows that the velocity at each point is perpendicular to the radius and equal to

$$v=\frac{1}{r}\frac{\partial\psi}{\partial\alpha}=\frac{\Gamma}{2\pi r}$$

The streamlines form circles. Let us find the velocity circulation along a circle. Since the length element is equal to $r\, d\alpha$ we can write for the circulation

$$\frac{\Gamma}{2\pi}\int_0^{2\pi}\frac{r\, d\alpha}{r}=\Gamma$$

Since curl $\mathbf{v}=0$ everywhere but the origin, the velocity has the same circulation around any closed path around the origin. The obtained pattern refers to a rectilinear filament vortex perpendicular to the flow plane. The existence of the vortex is seen from the fact that the circulation along a closed line is not zero.

Filament vortices need not be rectilinear. Since div curl $\mathbf{v}=0$, such lines either extend into infinity, close in on themselves, or terminate at a wall

or free surface. The velocity field produced by a filament vortex is the same as the magnetic field of a direct linear current (see Sec. 34) with corresponding boundary conditions.

The potential ψ of a filament vortex, unlike the source potential φ, is multiple-valued. In passing along a vortex filament the circulation Γ is added to the potential since α varies by 2π. Thus, $\oint d\psi = \Gamma$.

3. Investigate the flow pattern if the complex potential is given by the formula $w = \cosh^{-1} z = \psi + i\varphi = \cosh^{-1}(x + iy)$ (the units of measurement are so chosen that w and z are dimensionless).

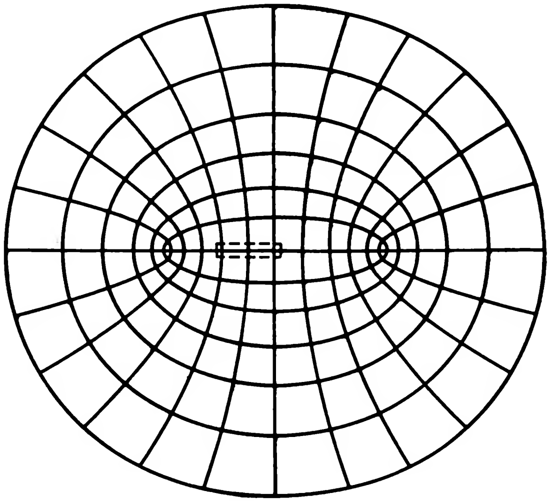


Figure 12

Solution. Obviously, $z = \cosh w = \cosh(\psi + i\varphi)$. We expand the hyperbolic cosine, where necessary go over to trigonometric functions and, after separating the real and imaginary parts, obtain:

$$x = \cosh\psi \cos\varphi, \qquad y = \sinh\psi \sin\varphi$$

In these equations, ψ and φ can be separated as follows:

$$\frac{x^2}{\cosh^2\psi} + \frac{y^2}{\sinh^2\psi} = 1, \qquad \frac{x^2}{\cos^2\varphi} - \frac{y^2}{\sin^2\varphi} = 1$$

It can be seen from this that, by virtue of the conformal nature or the mapping, the lines $\psi =$ constant form a family of confocal ellipses, and lines $\varphi =$ constant are a family of confocal hyperbolas perpendicular to the ellipses at their points of intersection (Figure 12)[3]. In the limit, at

[3] Figures 12 and 13 are taken from Maxwell's *Treatise on Electricity and Magnetism*, 3rd edition (1891), 2 vols., reprint by Dover, New York (1954).

$\psi = 0$, we obtain a segment of the x axis lying between 1 and -1 and joining the foci.

If the equipotential lines are given by the equation $\varphi =$ constant, the streamlines are closed: they encircle the interfocal segment. In passing along this segment the variable φ receives an increment 2π. This means that the potential is multiple-valued. To this corresponds a velocity circulation other than zero along an ellipse or any closed curve encompassing the segment $-1 \leqslant x \leqslant 1$. One can imagine the x,y-plane being crossed along this segment by a vortex restricted to a strip in a plane perpendicular to x, y. The linear density of the vortex is equal to the velocity discontinuity between the upper and lower ends of the segment, which follows directly from Stokes' theorem [11.19], as applied to the closed path shown by the dashed line.

The flow need not necessarily be around a linear segment. A portion of the same streamline pattern is obtained if some ellipse, denoted by a heavier line, is treated as a solid boundary around which the fluid circulates. In these conditions the circulation is not zero, while curl **v** vanishes everywhere. This can occur only in a multiply-connected region, which cannot be drawn into a point (because of the solid obstacle).

If we take ψ as the velocity potential, we obtain a different flow pattern: a flow from the upper half-space to the lower through an orifice in the shape of a strip lying within the limits $-1 \leqslant x \leqslant 1$ in a plane perpendicular to x, y. Actually eddies always form at the edges of an orifice, so that the velocity field obtained here is highly idealized. However, in the upper half-plane it describes the flow pattern fairly accurately, with the exception of the region along the wall.

4. Prove that if a complex potential depending on z is stated in implicit form by the equation $z = w + e^w$, the flow pattern corresponds to the one in Figure 13.

Solution. The heavy lines indicate the sections of the half-planes perpendicular to the flow plane and the y axis. This is a problem on the discharge of a liquid from a two-dimensional channel into an infinite flooded volume.

5. As is known, when water flows out of a bathtub, a hollow vortex funnel forms around the axis of which the liquid rotates. Determine the shape of the funnel.

Solution. Circulation occurs in a doubly-connected region; therefore curl **v** vanishes everywhere. The velocity potential satisfies the Laplace equation and is multiple-valued, as in Exercise 2. It can therefore also be taken in the form $\alpha(\Gamma/2\pi)$. Hence, the velocity of rotation of the liquid is

$$v = \frac{\Gamma}{2\pi r}$$

Since curl **v** is everywhere zero, the strong form of Bernoulli's theorem can be applied. Taking two points on the surface of the funnel, we find that

on the free surface of the liquid $p = 0$, and pressure or enthalpy are eliminated from the equation expressing the theorem. Far from the funnel axis we

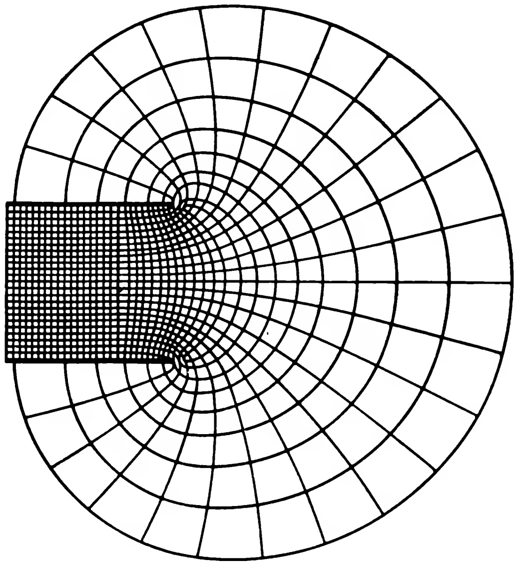


Figure 13

note that $U = 0$ and $v = 0$. At an arbitrary point, however, $U = gz$. Hence,

$$z = \left(\frac{\Gamma}{2\pi r}\right)^2 \frac{1}{g}$$

which expresses the relationship between the depth of the point and the radius of the funnel.

16

SOME PROBLEMS ON THE MOTION OF AN IDEAL FLUID

In this section we shall examine some problems on the isentropic motion of an ideal, incompressible fluid. But first we must establish a criterion of incompressibility, that is, the condition at which the density of a flowing fluid can be assumed constant.

Acoustic Waves. We shall proceed from the general equations (15.6) and (15.11), assuming the external force to be zero. Assuming

also that the change in density is small as compared with the undisturbed state, we put

$$\rho = \rho_0 + \rho' \tag{16.1}$$

where ρ_0 = constant, $\rho' \ll \rho_0$.

Now we replace the pressure gradient by the density gradient. Making use of the indicated inequality, we expand the pressure in a series, confining ourselves to the first term of the expansion:

$$p \approx p_0 + \left(\frac{\partial p}{\partial \rho}\right)_S \rho' \tag{16.2}$$

Here the derivative is taken at constant entropy. As p_0 = constant we replace grad p with grad ρ':

$$\text{grad } p = \left(\frac{\partial p}{\partial \rho}\right)_S \text{grad } \rho' \tag{16.3}$$

Let us also assume that the velocity of the fluid is small in the sense that expressions quadratic with respect to the velocity are small in comparison with linear terms. The products $\mathbf{v}$ times ρ' are also small. They should be treated as nonlinear terms in the equations. Thus Eqs. (15.6) and (15.11) are reduced to linear equations:

$$\frac{\partial \rho'}{\partial t} + \rho_0 \operatorname{div} \mathbf{v} = 0 \tag{16.4}$$

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -\left(\frac{\partial p}{\partial \rho}\right)_S \text{grad } \rho' \tag{16.5}$$

To eliminate $\mathbf{v}$ we must differentiate (16.4) with respect to time and take the divergence of (16.5). This yields an equation with respect to ρ':

$$\frac{\partial^2 \rho'}{\partial t^2} - \left(\frac{\partial p}{\partial \rho}\right)_S \nabla^2 \rho' = 0 \tag{16.6}$$

This equation has the form of the wave equation [18.5]. It describes the propagation of waves in a fluid with a velocity c, where

$$c^2 = \left(\frac{\partial p}{\partial \rho}\right)_S \tag{16.7}$$

From the thermodynamic inequalities (10.17), $(\partial p/\partial V)_S < 0$. But $dV = -d\rho/\rho^2$, so that (16.7) is always positive.

In Exercise 1 of Section 8 it was shown that isentropic derivatives are related to isothermal derivatives as follows:

$$\left(\frac{\partial p}{\partial \rho}\right)_S = \frac{c_p}{c_V}\left(\frac{\partial p}{\partial \rho}\right)_\theta \tag{16.8}$$

In an ideal gas $p = N\theta/(VM) = N\theta\rho/M$ (by (2.23) and (2.24)) (the ideal gas law). Therefore

$$\left(\frac{\partial p}{\partial \rho}\right)_\theta = \frac{N\theta}{M} = \frac{RT}{M} \tag{16.9}$$

where M is the molecular weight of the gas. The quantities $\frac{N}{M}$ and $\frac{R}{M}$ refer to one gram of the substance and not to one mole. The velocity of sound is comparable with the velocity of motion of individual molecules (cf. (2.14)).

It is now easy to establish the condition at which compressibility can be neglected in the motion of a liquid. The evaluation thus also holds for a gas. Let, for example, a solid body be moving in a fluid with a velocity $\mathbf{v}_0$. Going over to a frame of reference in which the body is at rest gives the case of the fluid flowing towards the body with the velocity $\mathbf{v}_0$. In this frame a particle of the fluid colliding "head on" with the body comes to a halt.

Apply to this particle the weak form of Bernoulli's theorem (15.17):

$$\frac{v_0^2}{2} + H_0 = H \tag{16.10}$$

Here, H_0 is the enthalpy of the flow far from the body, and H is the enthalpy of the halted particle of the liquid. Since we are concerned with the criterion of incompressibility, $(H - H_0)$ must be assumed small in comparison with H_0. Then (8.31) yields

$$H - H_0 = \frac{p - p_0}{\rho_0} = \frac{p'}{\rho_0} = \left(\frac{\partial p}{\partial \rho}\right)_S \frac{\rho'}{\rho_0} = c^2 \frac{\rho'}{\rho_0} \tag{16.11}$$

But from (15.18) it follows that $H - H_0 = v_0^2/2$, whence we obtain the criterion for satisfying the inequality $\rho' \ll \rho_0$:

$$\frac{\rho'}{\rho_0} = \frac{1}{2}\frac{v_0^2}{c^2} \ll 1 \tag{16.12}$$

The change in the velocity of any particle of the fluid must be small in comparison with the velocity of sound propagating through the fluid. The compressibility of the fluid matters only at transonic and supersonic flow velocities.

Taking the curl of Eq. (16.5), we find that for acoustic waves $\operatorname{curl} \mathbf{v} = 0$. Thus, for acoustic waves or other weak disturbances we can introduce the velocity potential according to the formula $\mathbf{v} = \operatorname{grad} \varphi$.

Surface Waves. Let us consider small oscillations of a free liquid surface. To them can be applied the linearized equations of hydrodynamics. However, unlike the case of the propagation of acoustic waves, the liquid should not be considered compressible, since the oscillations are reduced to form changes of the surface. We must con-

tinue to regard the quantity curl v equal to zero, since the term $(\mathbf{v}\cdot\nabla)\,\mathbf{v}$, which is quadratic with respect to velocity, is dropped, and in the right-hand side of the Euler equation of motion we have the gradient of the function $(U + p/\rho)$, where ρ is constant and U denotes the potential energy of the liquid in a gravitational field.

According to what has been said, at small disturbances the motion is potential, the potential in an incompressible fluid satisfying the Laplace equation $\nabla^2\varphi = 0$. For the sake of simplicity we shall treat

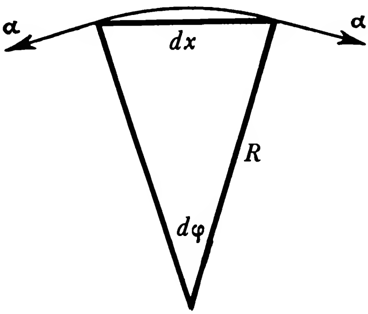


Figure 14

the motion as two-dimensional: the velocity field depends only upon the depth of a point below the z surface and upon the x coordinate, laid off along the undisturbed plane surface. Then the Laplace equation can be written as follows

$$\frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial z^2} = 0 \tag{16.13}$$

Now let us apply the Lagrange-Cauchy theorem (15.24) for a point lying on the surface of the liquid. We denote its vertical displacement by ζ. The square of the velocity of the point should be, as pointed out before, neglected. The potential energy per unit mass is, obviously, equal to gz, where g is the acceleration of free fall. For an ideal, incompressible liquid, instead of H we substitute p/ρ. Thus, Eq. (15.24) to the required approximation is written as follows:

$$\frac{\partial\varphi}{\partial t} + \frac{p}{\rho} + g\zeta = \text{constant}$$

where p is the pressure due to the warpage of the surface of the liquid during motion. As was pointed out in Section 14, the surface of a liquid is subject to surface tension α. In a plane surface the tension force has no component along the z axis. However, when the surface warps, such a component must appear. In that case it acts as external pressure.

Let us find it, making use of Figure 14, in which the section of curved surface is represented as a circular arc. As can be observed

from the drawing, the projection of the forces of surface tension on the vertical direction is equal to $-\alpha d\varphi$, that is, $-\alpha d\varphi/2$ on each side. Angle $d\varphi$ is equal to the ratio dx/R, where R is the radius of curvature of the surface. If $\zeta = \zeta(x, t)$ is the equation of the curve, we have from geometry the following approximate expression for the radius of curvature:

$$\frac{1}{R} = \left(\frac{\partial^2\zeta}{\partial x^2}\right)_t$$

Therefore, the projection of the resultant of the surface tension forces on the vertical direction is equal to

$$-\alpha\, dx \frac{\partial^2\zeta}{\partial x^2}$$

which implies that the pressure, that is, the force referred to unit surface, is given by the formula

$$p = -\alpha \frac{\partial^2\zeta}{\partial x^2} \tag{16.14}$$

In Figure 14 the surface curvature is negative so that p has the required sign.

Thus, at $z = 0$ Eq. (15.24) for a point on a free surface has the form

$$\frac{\partial\varphi}{\partial t} + g\zeta - \frac{\alpha}{\rho}\frac{\partial^2\zeta}{\partial x^2} = \text{constant} \tag{16.15}$$

We shall now look for the solution of (16.13) in the form of a harmonic wave travelling over the surface. As is known from [18.25], it must be a function of x and t according to the following law:

$$\varphi = \varphi_0(z)\cos(\omega t - kx) \tag{16.16}$$

Substituting this expression into (16.13), we find that $\varphi_0(z)$ satisfies the equation

$$\frac{d^2\varphi_0}{dz^2} - k^2\varphi_0 = 0 \tag{16.17}$$

If the depth of the liquid is great enough, only the solution of the form

$$\varphi_0 \sim e^{kz} \tag{16.18}$$

need be retained, since below the surface $z < 0$.

To satisfy condition (16.15) we differentiate it with respect to time and make use of the fact that on the surface

$$\frac{\partial\zeta}{\partial t} = v_z = \frac{\partial\varphi}{\partial z}$$

This yields

$$\frac{\partial^2\varphi}{\partial t^2} + g\frac{\partial\varphi}{\partial z} - \frac{\alpha}{\rho}\frac{\partial^2}{\partial x^2}\frac{\partial\varphi}{\partial z} = 0 \tag{16.19}$$

From (16.16) and (16.18), $\partial\varphi/\partial z = k\varphi$. From (16.16) we also find that $\partial^2\varphi/\partial t^2 = -\omega^2\varphi$ and $\partial^2\varphi/\partial x^2 = -k^2\varphi$. Cancelling out φ, we arrive at the equation expressing the dependence of frequency on the wave number k:

$$\omega^2 = kg + \frac{\alpha k^3}{\rho} \tag{16.20}$$

By analogy with electrodynamics such an equation is called a *dispersion equation* (see Sec. 37). The ratio of the frequency to the wave number is, according to [19.7], the phase velocity u of the wave. In accordance with (16.20) we can write:

$$u^2 = \frac{g}{k} + \frac{\alpha k}{\rho} \tag{16.21}$$

At small k's, that is, for long waves ($k = 2\pi/\lambda$), the first term (the term of gravitational origin) on the right predominates. Accordingly, long waves are called *gravitational.* At large k (short waves, or ripples) surface tension is more important. These waves are called *capillary.* The phase velocity has a minimum at

$$k_{\min} = \left(\frac{gp}{\alpha}\right)^{1/2} \tag{16.22}$$

The corresponding value of u is given by the formula

$$u_{\min} = \left(\frac{4\alpha g}{\rho}\right)^{1/4} \tag{16.23}$$

For water this amounts to approximately 26 cm-s^{-1}.

Oscillations of a Charged Drop. The theory of capillary waves proved extremely fruitful in its applications to the question of the stability of the atomic nucleus with respect to fission into two parts of more or less equal size. The interactions between particles in a nucleus take place over small distances (as between molecules), and they yield forces resembling surface tension in a liquid (Sec. 14). Surface tension in a nucleus opposes the forces of Coulomb repulsion between the protons of the nucleus. Coulomb forces are long-range forces. Therefore, the number of interacting protons increases as the square of the atomic number, Z^2. The average distance between two protons increases as $Z^{1/3}$, that is, as the size of the nucleus. It can be seen from this that the total Coulomb energy of the nucleus increases with the atomic number as $Z^{5/3}$. The surface energy increases as the square of the dimensions, that is, as $Z^{2/3}$.

If we assume that a nucleus has separated into two equal parts with atomic numbers $Z/2$ each, it is easy to see that for sufficiently great Z we gain in energy, and this additional energy can be realized in the form of the kinetic energy of the fragments scattered under

the action of the Coulomb repulsion forces. But before this can be realized, the nucleus has to be deformed somewhat and made into an elongated ellipsoid. Such an ellipsoid then continues to elongate by itself due to the action of the Coulomb forces.

The frequency of the oscillations of a spherical nucleus that draw it out into an ellipsoid is determined from a formula analogous to (16.20). It also has two terms on the right-hand side, but the first is of electrostatic rather than gravitational origin. Since the Coulomb force is directed outward from the surface, the first term in the frequency expression has a minus sign. There exists, therefore, a value of Z at which the nucleus becomes absolutely unstable with respect to infinitesimal distortions of the surface.

To find the general expression for the oscillation frequency of a charged particle, we must expand the oscillations not in travelling plane waves as in (16.16) but in standing spherical waves, using the Legendre polynomials [29.5]. Evaluations carried out by this method shortly after the discovery of uranium fission made it possible to obtain very significant semiquantitative results in fission theory (Niels Bohr and John A. Wheeler, J. I. Frenkel), which formed a basis for the subsequent development of nuclear physics.

Initially the theory of the fragmentation of charged (rain) drops was employed in atmospheric physics; it was decades later before it found another sphere of application, in nuclear physics.

Cavitation. Let us consider one more application of Eq. (15.24), this time in its exact form, taking into account quadratic terms. Let an evacuated spherical cavity of radius a have formed at some initial instant in an incompressible liquid at rest. We must establish the subsequent motion of the liquid, that is, the law of the collapsing of the cavity.

By the very formulation of the problem the motion is spherically symmetric, and the velocity has only a radial component. Then, from [11.46], the continuity equation is of the form

$$\frac{1}{r^2}\frac{\partial}{\partial r}(r^2 v)=0 \tag{16.24}$$

The solution of this equation, as is readily apparent, has the form

$$v=-\frac{A(t)}{r^2} \tag{16.25}$$

To this expression for the velocity corresponds the potential

$$\varphi=\frac{A(t)}{r} \tag{16.26}$$

We substitute the obtained results into Eq. (15.24), writing it in the form of an equation of the corresponding quantity at infinity

and on the surface of the cavity. At infinity the pressure is p_0, and on the cavity surface it is zero. We denote the current value of the cavity radius as r_0. Then we obtain the following form of Eq. (15.24):

$$\frac{p_0}{\rho}=\frac{1}{r_0}\frac{dA}{dt}+\frac{1}{2}\frac{A^2}{r_0^4} \tag{16.27}$$

Here, the left-hand side refers to $r=\infty$, and the right-hand side to $r=r_0\,(t)$. The equation for $r_0\,(t)$ is obtained from (16.25):

$$v\,(r_0)=\frac{dr_0}{dt}=-\frac{A\,(t)}{r_0^2} \tag{16.28}$$

The initial conditions for the obtained set of equations are: $r_0\,(0)=$ $=a$, $A\,(0)=0$, that is, at $t=0$ the liquid is at rest. Eliminating time from Eqs. (16.27) and (16.28), we obtain

$$\frac{dA}{dr_0}=\frac{1}{2}\frac{A}{r_0}-\frac{p_0r_0^3}{\rho A} \tag{16.29}$$

Multiplying by A reduces this equation to a linear equation with respect to A^2. It is also convenient to divide it by r_0 so as to simplify the dependence upon r_0. Taking r_0^2 as the new independent variable x and introducing the notation $A^2=y$, we have

$$\frac{dy}{dx}=\frac{1}{2}\frac{y}{x}-\frac{p_0}{\rho}x \tag{16.30}$$

The subsequent calculations are presented in Exercise 4; here we must say a few words concerning the significance of this problem. Cavities form in water in the rotation of propeller screws (cavitation). At the moment when the cavities collapse the liquid suddenly comes to a halt. But an instantaneous halt of any mass requires an infinitely large force. Actually, of course, compressibility has an effect, so that the pressure around the centre of collapse is not infinite but is simply very large. However, these small pressure peaks, transmitted through the liquid, combine to have a destructive effect on the screw blades. Special measures are taken to protect the blades from cavitation.

EXERCISES

1. Show that a travelling acoustic wave is polarized longitudinally.

Solution. The general expression for a travelling wave has the form (18.20):

$$\mathbf{v}=\mathbf{v}\left(t-\frac{\mathbf{rn}}{c}\right)$$

Since curl $\mathbf{v}=0$, $\dot{\mathbf{v}}\times\mathbf{n}=0$, or $\mathbf{v}\times\mathbf{n}=0$. This means that vectors $\mathbf{n}$ and $\mathbf{v}$ are colinear.

2. Find the dispersion law for surface waves in a reservoir of finite depth d.

Solution. If the solution of (16.17) is taken in the form

$$\varphi_0 = \cosh k(z + d)$$

then $v_z = k \sinh k(z + d)$ becomes zero at $z = -d$, as should be at a solid boundary. The dispersion law follows from the equation

$$\omega^2 = \left(kg + \frac{\alpha k^3}{\rho}\right) \tanh kd$$

3. Determine the minimum group velocity of propagation of surface waves.

Hint. Group velocity is equal to $d\omega/dk$ (see [19.8]).

4. Determine the time of collapse of a cavity in the cavitation problem.

Solution. The initial condition in Eq. (16.30) is that at $x = a^2$, $y = 0$. With this in mind, Eq. (16.30) has the following solution:

$$y = \frac{p_0}{\rho} x^{1/2} \int_x^{a^2} \frac{x\,dx}{x^{1/2}} = \frac{2}{3} \frac{p_0}{\rho} r_0 (a^3 - r_0^3)$$

whence

$$v(r_0) = \frac{dr_0}{dt} = -\frac{A(t)}{r_0^2} = -\left[\frac{2}{3} \frac{p_0}{\rho} \left(\frac{a^3}{r_0^3} - 1\right)\right]^{1/2}$$

The time of collapse is found as follows:

$$t = \int_0^a \frac{dr_0}{-v(r_0)} = a\left(\frac{3}{2} \frac{\rho}{p_0}\right)^{1/2} \int_0^1 \frac{x^{3/2}\,dx}{(1 - x^3)^{1/2}}$$

$$= a\left(\frac{\rho}{6p_0}\right)^{1/2} \int_0^1 z^{-1/6}(1 - z)^{-1/2}\,dz$$

The integral involved here is equal to 2.23.[4]

5. Prove that the oscillation period of a liquid in a U-shaped tube equals the oscillation period of a pendulum whose length is one-half the height of the liquid column.

---

[4] For the method of calculating this integral see G. A. Korn and T. M. Korn, *Mathematical Handbook for Scientists and Engineers*, 2nd ed., McGraw-Hill, New York, 1968, p. 823.

17

MECHANICS OF A VISCOUS INCOMPRESSIBLE FLUID

The Viscous Stress Tensor. If a fluid is at rest, its pressure is normal to any surface. This means that the nondiagonal components of the stress tensor are zero, and the diagonal components are equal to each other. In simpler terms this is to say that a fluid assumes the shape of the vessel into which it is poured: at rest it does not resist changes in form.

However, in the deformation of a fluid viscous forces come into play, provided the motion is at a finite velocity. Here we shall solve problems in which, unlike the problems on the flow of ideal fluids considered in Sections 15 and 16, a full picture of the flow pattern can be constructed only with due account of viscosity.

For a quantitative characterization of viscosity we must find the relationship between the tensor of the stresses within the liquid and the kinematic tensor describing the nonuniform velocity distribution. A spatially uniform velocity field cannot produce viscous stresses, since in such a field there is no displacement of the fluid particles relative to one another.

In order to determine the stress tensor over a non-uniform velocity field certain assumptions must be made. First of all, it is assumed that p_{ik} depends on the velocity distribution at a given instant and close to a given point in space. The value of p_{ik} is established in such a small time interval and in such a small length segment that no appreciable changes take place in the macroscopic motion of the fluid. Furthermore, it is assumed that the nonuniformity is small enough for the first derivatives of the velocities with respect to the coordinates, $\partial v_i/\partial x_k$, to describe it. Finally, it is assumed that these derivatives are small and terms quadratic with respect to them can be neglected. These conditions are usually satisfied in subsonic flow of water or air. At supersonic flow discontinuities that require a special description occur (see Sec. 24).

In order to express the symmetric tensor p_{ik} in terms of the tensor $\partial v_k/\partial x_i$, symmetric combinations of the components $\partial v_k/\partial x_i$ must be formed. Only two such expressions can be developed: $\delta_{ix}(\partial v_e/\partial x_e)$ and $(\partial v_k/\partial x_i)+(\partial v_i/\partial x_k)$. Hence, the dependence of the *viscous stress tensor* on the tensor $\partial v_k/\partial x_i$ has the following form:

$$p_{ik}=\eta\,(\partial v_k/\partial x_i+\partial v_i/\partial x_k)+\zeta'\delta_{ik}\frac{\partial v_e}{\partial x_e} \qquad (17.1a)$$

In writing Eq. (17.1*a*) it is more convenient to employ certain vector notations:

$$p_{ik} = \eta\left[\left(\frac{\partial v_k}{\partial x_i} + \frac{\partial v_i}{\partial x_k}\right) - \frac{2}{3}\delta_{ik}\,\mathrm{div}\,\mathbf{v}\right] + \zeta\delta_{ik}\,\mathrm{div}\,\mathbf{v} \qquad (17.1b)$$

where

$$\zeta = \zeta' + \frac{2}{3}\eta$$

The constants η and ζ are called the *first* and *second viscosity*, respectively.

If the flow pattern is that of an incompressible fluid, div $\mathbf{v} = 0$, and there remains simply

$$p_{ik} = \eta\left(\frac{\partial v_k}{\partial x_i} + \frac{\partial v_i}{\partial x_k}\right) \qquad (17.2)$$

It can be seen from this equation that viscosity is characterized by the stresses that develop when the layers of the fluid slip relative to one another. Thus, if the flow is parallel, that is, the velocity has only one component, say v_y, which depends on one coordinate perpendicular to the y axis, in accordance with (17.2) there appears a stress

$$p_{xy} = p_{yx} = \eta\frac{\partial v_y}{\partial x} \qquad (17.3)$$

Here the volume of the fluid does not change (div $\mathbf{v} = \partial v_y/\partial y = 0$). Thus, the first viscosity η belongs to the type of stress involved in changes in form of the fluid, as in the slipping of layers over one another. In most applications it is this viscosity that matters.

To establish the meaning of the second viscosity, ζ, consider uniform expansion of a fluid in all directions. The velocity at a given point in this case is proportional to the radius vector of that point: $\mathbf{v} = a\mathbf{r}$, $v_i = ax_i$, a = constant, and $\rho = \rho(t)$. The relative rate of change of the volume is div $\mathbf{v} = 3a$, as can be seen from the continuity equation (15.6):

$$-\frac{1}{\rho}\frac{\partial\rho}{\partial t} = \frac{1}{V}\frac{\partial V}{\partial t} = \mathrm{div}\,\mathbf{v} = 3a \qquad (17.4)$$

Thus, the density really depends only on time, and not on the coordinates, which corresponds to a uniform expansion process. The first term in (17.1*a*) vanishes, because the expansion law (17.4) yields $\partial v_i/\partial x_k = a\delta_{ik}$, leaving

$$p_{ik} = 3a\zeta\delta_{ik} = \frac{3}{V}\frac{\partial V}{\partial t}\delta_{ik} \qquad (17.5)$$

Thus, the second viscosity describes the strains that appear when the volume of a fluid changes. Note that the expression (17.5) has the Pascal form (15.7): p_{ik} is uniform pressure in all directions. But unlike static pressure it depends not on the volume itself but upon its rate of change.

In Section 8 it was shown that a process occurring in such conditions is irreversible, that is, it results in an increase in entropy. In a reversible process pressure depends only upon the volume at the given instant.

The second viscosity, ζ, becomes especially large when the process of establishing equilibrium in expansion or compression is in some way impeded. Suppose, for example, that the fluid is a gas whose molecules possess translational, rotational and vibrational degrees of freedom. In molecular collisions energy is easily transferred between the former two types of degrees of freedom. If the vibrational energy quanta of the molecules are considerably greater than the mean energy of thermal motion, θ, the molecules are excited in very few collisions. A highly improbable collision is required—with a molecule whose translational or rotational energy is, by some freak, substantially greater than θ—for a vibration quantum of one of the participants in the collision to be excited. In such conditions statistical equilibrium over the vibrational degrees of freedom is established slower than over the other degrees of freedom. In fast density fluctuations, acoustic, for example, equilibrium so to say "lags behind". In the absence of equilibrium there is always a finite rate of change towards equilibrium, that is, an irreversible process. The irreversible transfer of energy to internal degrees of freedom leads to the damping of orderly motions of the fluid (Exercise 1).

In air, where the vibrational quanta are large, at room temperatures there is not enough time for their excitation at all, and ζ is small. In carbon dioxide (CO_2) there exist deformation vibrations (Sec. 3) of comparatively low frequency. Therefore at room temperatures ζ in carbon dioxide is greater than in air.

The first viscosity, η, also causes irreversible processes in a fluid.

The Navier-Stokes Equations. We shall now write the equations of motion of a viscous fluid, using the general equation (15.5). For this we must express the stresses as a sum of the Pascal term referring to an ideal fluid and the stresses due to viscous forces:

$$\rho \frac{dv_i}{dt} = -\frac{\partial p}{\partial x_i} + \rho f_i + \eta \left(\frac{\partial^2 v_i}{\partial x_k^2} + \frac{\partial^2 v_k}{\partial x_i \, \partial x_k} - \frac{2}{3} \delta_{ik} \frac{\partial}{\partial x_k} \operatorname{div} \mathbf{v} \right)$$

$$+ \zeta \delta_{ik} \frac{\partial}{\partial x_k} \operatorname{div} \mathbf{v}$$

where the expression $\partial^2/\partial x_k^2$ involves summation with respect to the subscript k. But $\partial v_k/\partial x_k$ = div v. Consequently

$$\rho \frac{dv_i}{dt} = -\frac{\partial p}{\partial x_i} + \rho f_i + \eta \nabla^2 v_i$$
$$+ \eta \frac{\partial}{\partial x_i} \operatorname{div} \mathbf{v} - \frac{2}{3} \eta \frac{\partial}{\partial x_i} \operatorname{div} \mathbf{v} + \zeta \frac{\partial}{\partial x_i} \operatorname{div} \mathbf{v}$$

Going completely over to vector notation, we obtain the general equations of motion for a viscous fluid (the *Navier-Stokes equations*):

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\, \mathbf{v} \right) = -\operatorname{grad} p + \rho \mathbf{f} + \eta\, \nabla^2 \mathbf{v}$$
$$+ \left(\zeta + \frac{\eta}{3} \right) \operatorname{grad} \operatorname{div} \mathbf{v} \qquad (17.6a)$$

For an incompressible fluid they are simplified thus:

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\, \mathbf{v} \right) = -\operatorname{grad} p + \rho \mathbf{f} + \eta\, \nabla^2 \mathbf{v} \qquad (17.6b)$$

Together with the continuity equation (div v = 0) they form a complete set. Note also that in going over to curvilinear coordinates $\nabla^2 \mathbf{v}$ is conveniently written as grad div v — curl curl v. The left-hand side of the equation also changes in the passage to curvilinear coordinates, because the very concept of a vector in curvilinear coordinates is different than in rectilinear. For example, the motion of a point along a coordinate line in curvilinear coordinates is nonrectilinear and cannot be treated as free.

As compared with (15.11) (the Euler equation of motion), the Navier-Stokes equations (17.6*b*) are of a higher order of derivatives with respect to the coordinates. This means that a solution to (17.6*b*) requires additional boundary conditions.

Experience reveals that a viscous fluid does not slip along a wall. The velocity of flow at a motionless wall is zero; at a moving wall it is equal to the velocity of the wall. In an ideal fluid the boundary condition was superimposed only upon the normal velocity component on the wall; in a viscous fluid the boundary condition is superimposed also upon the tangential velocity component. It can thus be seen that if there are solid walls, a smooth transition from a viscous to an ideal fluid over the whole flow region is impossible. At the wall there is always a layer in which viscous forces are essential.

Liberation of Heat in Viscous Flow of a Fluid. We shall now show how mechanical energy transforms into heat in viscous flow. For simplicity we shall restrict ourselves to the case of an incompressible fluid.

Consider Eq. (15.13). In the case of a viscous fluid, in the right-hand side instead of zero there will be the product $\eta \mathbf{v} \nabla^2 \mathbf{v}$. Therefore, the energy balance equation will include losses. We rewrite it[5] in tensor notation:

$$\frac{\partial}{\partial t} \rho \left(\frac{v^2}{2} + U \right) + \frac{\partial}{\partial x_k} \rho v_k \left(\frac{v^2}{2} + \frac{p}{\rho} + U \right) = \eta v_i \frac{\partial^2 v_i}{\partial x_k^2} \quad (17.7)$$

The expression involving viscosity must be transformed by parts:

$$v_i \frac{\partial^2 v_i}{\partial x_k} = \frac{\partial}{\partial x_k} v_i \frac{\partial v_i}{\partial x_k} - \left(\frac{\partial v_i}{\partial x_k} \right)^2$$

Now we integrate Eq. (17.7) over the whole volume of the fluid. Then in front of the whole expression, the terms involving $\partial/\partial x_k$ will reduce to integrals over the surface and become zero. In the left-hand side of the equation this is so because the energy does not dissipate outside the fluid (specific heat is not considered), and in the right-hand side this is due to the conditions imposed on the velocity. On solid walls the velocity is zero, while on a free surface there can be no stresses, that is, the quantity $\partial v_k/\partial x_i$, which is proportional to stress, vanishes.

Consequently the following terms remain in the energy balance equation:

$$\frac{\partial}{\partial t} \int \rho \left(\frac{v^2}{2} + U \right) dV = -\eta \int \left(\frac{\partial v_i}{\partial x_k} \right)^2 dV \quad (17.8)$$

According to the second law of thermodynamics the mechanical energy of nonrandom motion can only decrease and is transferred to the internal energy of the fluid. Consequently, the viscosity η is always greater than zero.

Equation (17.8) describes the continuous transfer of the energy of mechanical motion into the energy of molecular motion, or the dissipation of energy, as it is called. That is why the Navier-Stokes equations refer to nonconservative motion involving friction [Sec. 1], in this case, viscous.

Reynolds Number. Suppose that no external forces $\mathbf{f}$ are acting on an incompressible fluid. Transfer grad p to the left-hand side of the Navier-Stokes equation, leaving the expression involving density on the right:

$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right) + \nabla p = \eta \frac{\partial^2 \mathbf{v}}{\partial x_k^2} \quad (17.9)$$

---

[5] The method of deriving (15.13) shows that in the case of an incompressible fluid the internal energy E is not involved in the energy density at all, and only p/ρ remains in place of the enthalpy H.

Now consider a problem, say of steady-state flow around a body. If the shape of the body is given, it is characterized by one linear pimension l. We shall measure length in units of l and accordingly prime the coordinate x/l and the gradient. Since by definition the problem is that of a steady-state flow, we discard $\partial \mathbf{v}/\partial t$. Then Eq. (17.9) takes the form

$$\frac{1}{l}\rho(\mathbf{v}\cdot\nabla')\,\mathbf{v}+\frac{1}{l}\,\mathrm{grad}'\,p=\frac{\eta}{l^2}\,\frac{\partial^2\mathbf{v}}{\partial x_k'^2} \tag{17.10}$$

Denote the velocity of the incident flow at a great distance from the body by v, and divide Eq. (17.10) by v. Furthermore, denote by $\mathbf{v}'$ the velocity measured in units of v. Then the Navier-Stokes equations take the form

$$(\mathbf{v}'\cdot\nabla')\,\mathbf{v}'+\frac{\mathrm{grad}'\,p}{\rho v^2}=\frac{\eta}{\rho v l}\,\frac{\partial^2\mathbf{v}'}{\partial x_k'^2} \tag{17.11}$$

The quantity

$$N_{\mathrm{Re}}\equiv\frac{\rho v l}{\eta} \tag{17.12}$$

is called the *Reynolds number*. It is the only dimensionless parameter in Eq. (17.11). If the equation is solved for one specific problem, we thereby obtain a solution for flow around all bodies of similar shape, provided the number N_{Re} is the same for all cases. The answer is in the corresponding units: length in the dimension of l, velocity in the dimension of v', and pressure in the dimension of pv^2. Instead of an analytical solution we may make use of measurements carried out on models, provided the Reynolds number in the flow was the same as in the real problem.

Using ideas of similarity, we can determine the manner in which a force F acting on a body depends upon the dimensions of the body and the flow parameters. The combination $\rho v^2 l^2$ has the dimension of force. In the expression for force it must be accompanied by a coefficient depending only upon the Reynolds number:

$$F=\rho v^2 l^2 f(N_{\mathrm{Re}}) \tag{17.13}$$

A fluid is characterized not by its viscosity and density taken separately, but by their ratio

$$\nu=\frac{\eta}{\rho} \tag{17.14}$$

This quantity is called *kinematic viscosity*, because its dimensions do not include the unit of mass: $[\nu]=\mathrm{cm^2 s^{-1}}$.

The Reynolds number characterizes the relative values of the inertia and viscosity terms in the Navier-Stokes equations. High values of N_{Re} indicate a predominance of inertia terms, that is, $\rho(\mathbf{v}\cdot\nabla)\,\mathbf{v}$; small N_{Re} indicates a predominance of viscosity terms,

that is, $\eta\nabla^2\mathbf{v}$. In very slow motion, when $N_{\mathrm{Re}} \ll 1$, the role of inertia terms is very small. In the opposite limiting case, when $N_{\mathrm{Re}} \gg 1$, the flow is almost that of an ideal fluid. But as said before, close to a solid wall there is always a flow region where viscosity has a marked effect. In the limit at $N_{\mathrm{Re}} \gg 1$, these regions may become very small, but they never disappear altogether.

Viscous Flow in a Cylindrical Pipe. Consider a problem in which the inertia terms are negligible not because of their small values but simply as a consequence of symmetrical flow. If an incompressible fluid is flowing through a cylindrical pipe, the velocity is directed along the z axis of the pipe and depends only upon the distance r to the axis. The operator $(\mathbf{v}\cdot\nabla)$ is reduced to $v\,(d/dz)$, and when applied to $v\,(r)$ yields zero. There is no derivative with respect to time, since the flow is steady. In that case the Navier-Stokes equations take the form

$$\eta\,\nabla^2 v = -\frac{\partial p}{\partial z} \tag{17.15}$$

To satisfy this equation we must assume the pressure p to be a function only of z, and a linear function at that. Indeed, if p were a function of the radius, there would be a radial pressure gradient and consequently a radial velocity component, which would contradict our assumption. Furthermore, for the left-hand side of (17.15) to be independent of z, the derivative $\partial p/\partial z$ must be a constant quantity. These assumptions make it possible to satisfy all the conditions of the problem.

Introducing the notation $\partial p/\partial z \equiv p'$ and employing cylindrical coordinates, we obtain

$$\frac{1}{r}\frac{d}{dr}r\frac{dv}{dr} = \frac{p'}{\eta} \tag{17.16}$$

We also note that the velocity cannot become infinite at the pipe axis or zero at the wall (at $r = a$). The solution of (17.16) in these conditions has the form

$$v = -\frac{p'}{\eta}\int_0^a \frac{dr}{r}\int_0^r r\,dr = -\frac{p'}{4\eta}(a^2 - r^2) \tag{17.17}$$

The minus sign indicates that the liquid is flowing in the direction of lower pressure, against the gradient of p. The total discharge of the fluid is

$$2\pi\rho\int_0^a vr\,dr = -\frac{\pi}{8}\frac{\rho p'}{\eta}a^4 = \frac{\pi}{8}\frac{p_1 - p_2}{l\nu}a^4 \tag{17.18}$$

(Poiseuille's formula). Here, $p_1 - p_2$ is the pressure difference between the pipe entrance and exit, l the pipe length, and ν the kinematic viscosity. We have thus obtained an exact solution of the Navier-Stockes equations.

Uniform Fall of a Solid Sphere in a Viscous Fluid. When a small sphere falls in a viscous fluid, at some velocity the resisting force F balances the force of gravity. In these conditions the velocity becomes constant and depends upon the radius of the sphere, the force of gravity acting upon it, and the viscosity and density of the fluid. If the sphere is small, its velocity of uniform motion and, hence, the velocity of the fluid flow around it, are small. So is the Reynolds number that characterizes that flow, so that the inertia terms in the Navier-Stokes equations can be neglected. But the inertia terms introduce density into hydrodynamic equations. Since in the present problem of slow fall of a sphere we neglect these terms, the force must be expressed only in terms of the viscosity, the velocity of the sphere, and its radius. Dimensional analysis immediately reveals that there is only one such expression, which up to a numerical coefficient is

$$F \sim \eta v a$$

Calculations reveal that the coefficient of proportionality is equal to 6π. Hence

$$F = 6\pi\eta v a \tag{17.19}$$

This is *Stokes' law.*

The work done by the force of gravity in unit time in the fall of a sphere in a fluid is $Fv_0 = 6\pi\eta R v_0^2$. It results in heating of the fluid, since in steady motion the kinetic energy of the system does not increase (this holds true for a body of any shape in steady motion in a viscous fluid).

Mobility. In the case of sufficiently slow uniform motion of a a body, formula (17.19) can be written in the form

$$v = \tilde{\omega} F \tag{17.20}$$

where the coefficient $\tilde{\omega}$ is equal to $(6\pi\eta R)^{-1}$. It is termed the *mobility* of the body in the fluid.

Let us now suppose that we have a fluid with particles suspended in it on which no forces act—neither viscous, nor hydrostatic, nor gravity. But the particles are in some way nonuniformly distributed throughout the fluid volume. Such a state of the system cannot be the most probable one. As they wander through the fluid under the action of random pressure fluctuations (Brownian motion), the par-

ticles will on the average be distributed uniformly throughout the volume. If their concentration n varies, they will pass, on the average, more frequently from places of greater concentration to places of smaller concentration. If the concentration gradients are not too great, the flow of particles per unit time across unit area is proportional to the concentration gradient:

$$\mathbf{j} = -D \operatorname{grad} n \tag{17.21}$$

The factor D is called the *diffusivity* of the particles in the fluid. The minus indicates that diffusion occurs from greater to lesser concentrations, that is, in the opposite direction of grad n.

Now let the particles be subject to the action of a force field $\mathbf{f}$, which we shall consider conservative:

$$\mathbf{f} = -\operatorname{grad} U$$

A steady distribution of the particles will be established if the sum of the flow, nv, due to the mobility of the particles, and of the diffusion flow, vanishes:

$$-\tilde{\omega} n \operatorname{grad} U - D \operatorname{grad} n = 0 \tag{17.22}$$

or

$$-\tilde{\omega} \operatorname{grad} U - D \operatorname{grad} \ln n = 0 \tag{17.23}$$

The solution of this equation has the form:

$$n = n_0 e^{-\tilde{\omega} U/D} \tag{17.24}$$

But in equilibrium the particles are subjected to the Boltzmann distribution (2.38):

$$n = n_0 e^{-U/\theta} \tag{17.25}$$

In deducing the Boltzmann law no restrictions were imposed on the size of the particles. The only condition was that their equilibrium distribution be statistical. Comparing (17.24) with (17.25), we come to the conclusion that

$$D = \tilde{\omega}\theta \tag{17.26}$$

This equation is called the *Einstein relationship between mobility and diffusivity*. If the mobility of the particles is known, this gives their diffusivity, and vice versa.

Note that Eq. (17.26), like the concept of diffusion itself, is applicable not only to particles of macroscopic dimensions. It also holds, for example, for ions in an electrolyte and for atomic impurities in a substance.

EXERCISES

1. A plane monochromatic acoustic wave of frequency ω propagates through a viscous medium. Determine its attenuation on unit length.

Solution. Using (16.5), (16.7), and (17.6), we can write the equations for acoustic disturbances:

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} = -c^2 \operatorname{grad} p' + \eta \nabla^2 \mathbf{v} + \left(\zeta + \frac{\eta}{3}\right) \operatorname{grad} \operatorname{div} \mathbf{v}$$

$$\frac{\partial \rho'}{\partial t} = -\rho_0 \operatorname{div} \mathbf{v}$$

whence

$$\frac{\partial^2 \rho}{\partial t^2} = c^2 \nabla^2 \rho' - \left(\zeta + \frac{4}{3}\eta\right) \nabla^2 \frac{1}{\rho_0} \frac{\partial \rho'}{\partial t}$$

If

$$\rho' = \rho'_a e^{-i\omega t + ikx}$$

(the x axis is directed along $\mathbf{k}$), then

$$\omega^2 = c^2 k^2 + \frac{i\omega k^2}{\rho_0}\left(\zeta + \frac{4}{3}\eta\right)$$

Consequently

$$k^2 = \frac{\omega^2}{c^2}\left[1 + \frac{i\omega}{\rho_0 c^2}\left(\zeta + \frac{4}{3}\eta\right)\right]^{-}$$

or approximately

$$k = \frac{\omega}{c}\left[1 - \frac{i\omega}{2\rho_0 c^2}\left(\zeta + \frac{4}{3}\eta\right)\right]$$

The imaginary part of k yields the required attenuation, which occurs according to the exponential law

$$\rho' = \rho'_a \exp\left[-i\omega t + \frac{i\omega x}{c} - \frac{x\omega^2}{2\rho_0 c^3}\left(\zeta + \frac{4}{3}\eta\right)\right]$$

If the second viscosity, due to irreversible processes accompanying volume change, is not too large, then in gases attenuation of sound of the same order of magnitude occurs by virtue of heat transfer from compressed sections to rarified sections. The physical reason for the coincidence of the orders of magnitude consists in that viscosity and heat conduction are due to similar mechanisms of molecular collision.

2. A viscous incompressible fluid is placed between two rotating infinite coaxial cylinders whose angular velocities and radii are, respectively, ω_1 and ω_2, and r_1 and r_2. Determine the velocity field for time-constant ω_1 and ω_2.

Solution. Going over to polar coordinates r, φ, z, we see that the velocity has only the component $v_\varphi = v$, which is a function of r. Like the problem

on fluid motion through a pipe, the inertia terms here vanish. For the Laplacian of $\nabla^2 \mathbf{v}$ we have

$$(\nabla^2 \mathbf{v})_\varphi = -\mathrm{curl}_\varphi \, \mathrm{curl}\, \mathbf{v} = \frac{d}{dr}\left(\frac{1}{r}\right)\frac{dr}{dv}\, rv = 0$$

whence

$$v = Cr + \frac{C_1}{r}$$

Only the pressure gradient and the centrifugal term $\rho v^2/r$ have radial components in the Navier-Stokes equations. It is therefore easy to determine the dependence of the pressure upon the radius if we know the dependence of the velocity on the radius.

At the walls the conditions $v(r_1) = \omega_1 r_1$ and $v(r_2) = \omega_2 r_2$ hold. Therefore

$$v = \frac{\omega_2 r_2^2 - \omega_1 r_1^2}{r_1^2 - r_2^2}\, r + \frac{(\omega_1 - \omega_2)\, r_1^2 r_2^2}{(r_1^2 - r_2^2)\, r}$$

3. At an initial instant, N particles capable of diffusing in a given medium are assembled within a very small volume. Determine their distribution at subsequent instants of time.

Solution. The density of the particles depends only on the distance r, so that the diffusion flux is radial. Then

$$j = -D\frac{\partial n}{\partial r}$$

The law of conservation of the number of particles has the form

$$\frac{\partial n}{\partial t} = -\mathrm{div}\, \mathbf{j} = \frac{D}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial n}{\partial r}$$

Since the initial volume in which the particles were contained was assumed infinitesimal, the problem does not involve characteristic quantities with the dimensions of length. Therefore, r can be involved only in the dimensionless combination $\xi = r^2/(Dt)$, where $[D] = [l^2/t]$. For the conservation of the total number of particles the coefficient of the function involving ξ must be proportional to $(Dt)^{-3/2}$. Indeed, then and only then is the number of particles independent of time:

$$N = 4\pi \int_0^\infty nr^2\, dr = 4\pi \int_0^\infty f\left(\frac{r^2}{Dt}\right) \frac{r^2\, dr}{(Dt)^{3/2}} = 4\pi \int_0^\infty f(\xi)\, \xi\, d\xi^{1/2}$$

From the expression

$$n = (Dt)^{-3/2} f\left(\frac{r^2}{Dt}\right)$$

we find the derivatives

$$\frac{\partial n}{\partial t}=-\frac{3}{2}\frac{f}{t\,(Dt)^{3/2}}-\frac{1}{t\,(Dt)^{3/2}}\xi\frac{df}{d\xi}$$

$$=-\frac{1}{t\,(Dt)^{3/2}}\left(\frac{3}{2}f+\xi\frac{df}{d\xi}\right)$$

$$\frac{\partial n}{\partial r}=\frac{2r}{(Dt)^{5/2}}\frac{df}{d\xi},\qquad \frac{D}{r^2}\frac{\partial}{\partial r}r^2\frac{\partial n}{\partial r}=\frac{1}{t\,(Dt)^{3/2}}\left(6\frac{df}{d\xi}+4\xi\frac{d^2f}{d\xi^2}\right)$$

Substituting these into the equation for n, we find a differential equation for f:

$$4\xi\ddot{f}+6\dot{f}+\xi\dot{f}+\frac{3}{2}f=0$$

where

$$\dot{f}\equiv\frac{df}{d\xi}$$

The function

$$f=Ce^{-\xi/4}$$

is a solution to the equation. (The second solution is not damped at infinity and thus does not assure the constancy of N, since the integral $\int\limits_0^\infty n\xi^{3/2}\,d\xi$ diverges.) From the conservation of N we obtain the constant C:

$$N=4\pi C\int\limits_0^\infty e^{-\xi/4}\frac{1}{2}\xi^{1/2}\,d\xi=2^4\pi C\int\limits_0^\infty e^{-\zeta}\zeta^{1/2}\,d\zeta=8\pi^{3/2}C$$

$$C=\frac{N}{8\pi^{3/2}}$$

Consequently

$$f=\frac{N}{8\pi^{3/2}}e^{-\xi/4}$$

and

$$n=\frac{N}{(4\pi Dt)^{3/2}}e^{-r^2/(4Dt)}$$

As $t\to 0$, this function tends to zero everywhere except the origin of the coordinate system and therefore satisfies the initial condition. The obtained formula provides one of the representations of the three-dimensional δ function [Sec. 26].

If the particles were initially at a point $\mathbf{r}=\mathbf{r}'$, that point can be taken as the origin of the coordinate system. In that case the solution has the form

$$n\,(\mathbf{r},\ t,\ \mathbf{r}')=\frac{N}{(4\pi Dt)^{3/2}}\exp\left[-\frac{(\mathbf{r}-\mathbf{r}')^2}{4Dt}\right]$$

As $t\to 0$, it becomes $\delta\,(\mathbf{r}-\mathbf{r}')$.

4. Show that $\overline{r^2} = 6Dt$, that is, that a diffusing particle recedes from its initial position at a rate proportional to $t^{1/2}$.

5. Find the function $n(\mathbf{r}, t)$ satisfying the diffusion equation for an arbitrary initial spatial distribution of particles $n(\mathbf{r}, 0) = n_0(\mathbf{r})$.

Solution. Since the diffusion equation is linear, the solution has the form

$$n(\mathbf{r}, t) = \int \frac{n(\mathbf{r}')}{(4\pi Dt)^{3/2}} \exp\left[-\frac{(\mathbf{r}-\mathbf{r}')^2}{4Dt}\right] dV'$$

The operator $\left(\frac{\partial}{\partial t} - \frac{D}{r^2}\frac{\partial}{\partial r} r^2 \frac{\partial}{\partial r}\right)$ applied to both parts of this equation is permutable with the integration with respect to dV' and yields zero. The initial condition is satisfied because

$$n(\mathbf{r}, 0) = \int n_0(\mathbf{r}')\, \delta(\mathbf{r}-\mathbf{r}')\, dV' = n_0(\mathbf{r})$$

The function used to find the solution for an arbitrary initial condition from the solution for a point initial distribution is called the *Green function* of the initial equation. Account here is also taken of the boundary condition $n(r \to \infty) = 0$.

18

MOTION OF BODIES IN AN INCOMPRESSIBLE FLUID

Laminar and Turbulent Flow. The type of flow dealt with up till now is called *laminar*. The fluid flows as it were in stratified layers which do not mix. Such flow is observed at relatively low Reynolds number: in the tens or less. At high Reynolds numbers, as experience shows, the flow is usually highly irregular (fluctuating) and accompanied by extensive mixing.

In this connection there arises the question of the stability of laminar flow. Not every type of motion compatible with the equations of hydrodynamics necessarily occurs in fact. In some cases an infinitesimal perturbation in the initial condition with time considerably deflects the fluid from the given motion corresponding to the laminar solution of hydrodynamic equations. Similarly, unstable equilibrium is disturbed by an infinitesimal perturbing force.

Unstable motion in particle mechanics is also analogous to unstable equilibrium. An example is the motion of a point along the upper generatrix of a horizontal cylinder in a gravitational field. The slightest deviation from the generatrix causes the point to slide off the

cylinder. If a groove is made along the generatrix, a finite deflection from the path is needed for the point to be displaced a long way from it. This is an example of instability with respect to finite perturbations.

Friction forces usually stabilize the motion, but sometimes they destabilize it. A general investigation of the stability of motion is very complicated. As applied to fluids, it has been undertaken in few cases, and the results cannot, evidently, always be guessed intuitively, even in the qualitative aspect.

Of great use here is experimental research, which, most generally speaking, shows the following. When the Reynolds number characterizing laminar flow becomes greater than a certain critical value, the velocity begins to fluctuate, at first regularly. But when the Reynolds number becomes considerably greater than the critical value, the fluctuations become irregular in character, while remaining nevertheless steady. By the definition of steady state [17.5] this means that the time-average of the total derivative with respect to time of any quantity characterizing flow is equal to zero:

$$\overline{\frac{df}{dt}} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_{-\tau/2}^{\tau/2} \frac{df}{dt}\, dt$$

$$= \lim_{\tau\to\infty} \left\{ \frac{1}{\tau} \left[f\left(\frac{\tau}{2}\right) - f\left(-\frac{\tau}{2}\right) \right] \right\} = 0$$

Taking certain mathematical precautions, such motion can be vizualized as a combination of periodic motions, of fluctuations. Fluctuations are due to random causes affecting unstable laminar flow. Therefore individual fluctuations are characterized by an assembly of random, chaotically distributed, phases.

In this case fluid motion can to a considerable degree be investigated by the same methods as the state of a statistical system. Any classical statistical system requires for detailed description a vast assembly of initial conditions of motion of individual particles. Since this is not only impossible but also in practice unnecessary, probability description is used; with its help the mean values of various quantities can be determined. This makes it possible to investigate chaotic (turbulent) fluid motion. A set of a large number of initial phases is treated as a statistical assembly. In this case even the smallest fluctuations lie wholly in the domain of applicability of the Navier-Stokes equations, so that the atomic structure of a fluid has no relation to the problem of turbulence, provided the scale of the fluctuations is much greater than the mean free path of the molecules.

Obviously, not all the conclusions of statistical mechanics can be applied to the theory of turbulence. A statistical assembly is treated as a closed or quasi-closed system in which energy is exactly

or approximately conserved. In turbulent flow, on the other hand, according to the Navier-Stokes equations, there are always viscous forces, which lead to the dissipation of the energy of macroscopic motion. Therefore we cannot speak of any similarity to statistical equilibrium in turbulent flow. All that can be obtained is a quasi-steady state with constant mean energy dissipation at every point of space.

Boundary Layer. The Navier-Stokes thermodynamic equations are of a higher order than the Euler equations of motion, which do not take viscosity into account. Accordingly, an additional condition is imposed on the velocity of viscous flow: at the surface of a stationary solid body it must vanish. If the body is in motion, the velocity of the fluid at every point of the body is, as stated before, equal to the velocity of that point. In an ideal fluid it was sufficient only for there to be no fluid motion through the surface of the body.

It is obvious therefore that the limiting transition to the case of infinitesimal viscosity in the Navier-Stokes equations is by no means a simple operation. One cannot directly assume viscosity to be zero throughout the whole volume of the fluid. There are always viscous forces in a narrow (boundary) layer along a wall, and the motion of the fluid can be described only by the exact Navier-Stokes equations, which assure satisfaction of the boundary condition $\mathbf{v} = 0$ at the wall. The less the viscosity the thinner the layer. Its properties to a considerable degree determine the force exerted on a moving body by the fluid through which it is moving.

Drag and Lift. If a body is in uniform motion, it is more convenient to adopt a reference system in which it is at rest while the fluid is flowing around it in the opposite direction. At an infinite distance in front of the body the velocity of the fluid is constant and equal to the velocity of the body taken with the opposite sign. Close to the body, of course, the velocity does not form a constant field. The force with which the fluid acts upon the body must be calculated according to the velocity distribution.

In the most general case the force has three components in a coordinate system with two planes parallel to the main, or undisturbed, stream. But if the body is symmetrical with respect to one of these planes there are only two component forces. One of them is directed in the opposite direction of the body's velocity, and is called the *resisting force*, or *drag*. The other force is perpendicular to it and is called *lift*. In the case of a nonsymmetrical body there is also a deflecting force in the third, perpendicular, direction, but we shall not consider it here.

The problem is to find, as far as possible, the general expressions for the forces acting upon a body in a fluid.

Boundary Layer Separation. The first thing is to gain an understanding of the origin of these forces. For this let us examine flow patterns around bodies of different cross section.

If a body has abrupt edges in the cross section, as in Figure 15, the boundary layer of the fluid flowing around it is separated from the body and is carried away a considerable distance from it, forming the so-called wake.

In a viscous layer at the surface of the body curl **v** is not zero, since there exists a transverse velocity gradient. At a distance from

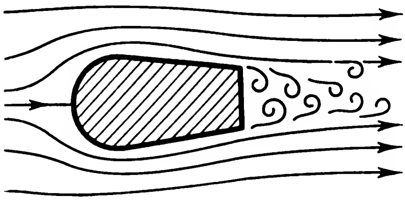


Figure 15

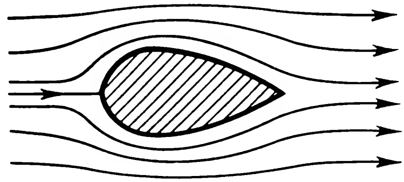


Figure 16

the wall, where the role of viscosity is not great, curl **v** attenuates very slowly, in accordance with the theorem on the conservation of velocity circulation.

As a consequence, a very long eddying wake forms behind the body, which, as will be shown, contributes to the drag in flow around not very thin airfoil-type cross sections. We shall consider airfoils separately.

The boundary layer may also separate from smoothly tapering cross sections.

Consider a smooth cross section of the type in Figure 16. At its widest point the streamlines are compressed. Consequently the velocity is highest. Then they begin to spread, and the velocity drops.

From Bernoulli's theorem (15.17), where velocity decreases pressure increases. This refers to an incompressible fluid, for which instead of enthalpy we should write p/ρ.

Let us apply Bernoulli's theorem to a section of the flow between points x and $x + \Delta x$:

$$\frac{1}{2}\left[v_0^2(x+\Delta x) - v_0^2(x)\right] = -\frac{1}{\rho}\left[p(x+\Delta x) - p(x)\right] \qquad (18.1)$$

Here subscript 0 is meant to indicate that the velocity's value is taken in the mainstream and not in the boundary layer.

The pressure does not change in the direction perpendicular to the velocity, since in steady flow $(\mathbf{v}\cdot\nabla)\,\mathbf{v} = -(1/\rho)\operatorname{grad} p$. It

follows from this that in the boundary layer the purely hydrostatic pressure gradient for p is the same as in the mainstream. As for the term $\nu(\partial^2 v_x/\partial y^2)$, due to viscosity, it can tend only to a finite limit because the thickness of the layer, y, as theory indicates, is proportional to $\nu^{1/2}$. Consequently, at sufficiently large pressure gradients $|(1/\rho)\,\mathrm{grad}\,p| \gg |\nu(\partial^2 v_x/\partial y^2)|$, at which Bernoulli's theorem is applicable (approximately), in the boundary layer we have the relationship

$$\frac{1}{2}[v^2(x+\Delta x) - v^2(x)] = -\frac{1}{\rho}[p(x+\Delta x) - p(x)] \tag{18.2}$$

Now, using (18.1), we obtain

$$v^2(x+\Delta x) - v^2(x) = v_0^2(x+\Delta x) - v_0^2(x) \tag{18.3}$$

whence it follows that

$$v^2(x+\Delta x) = v^2(x) + [v_0^2(x+\Delta x) - v_0^2(x)] \tag{18.4}$$

The expression in brackets on the right-hand side is negative since it refers to the flow around the tapering part of the body, while the point with the coordinate $x + \Delta x$ is downstream with respect to the point with coordinate x. If the negative quantity exceeds the positive term $v^2(x)$, then Eq. (18.4) cannot be satisfied since the quantity on the left is essentially positive. Consequently the boundary layer of the fluid can no longer continue to flow around the body. It must separate and transform into a wake.

This reasoning, based as it is on estimates, is to a considerable degree suggestive and is not proof that separation necessarily occurs. In real conditions separation occurs if the cross section does not taper smoothly enough.

At large Reynolds numbers the wake may also be turbulent. If turbulence is caused by the moving body itself, while the initial flow is laminar, the velocity fluctuations in the wake gradually attenuate, and at a great distance behind the body the flow in the wake also becomes laminar.

Expression of the Force Acting on a Body in Terms of the Velocity Distribution in the Wake. Let us now see how the velocity distribution in the wake is related to the force acting on the body. We shall proceed from Eq. (15.15), from which the tensor of the momentum flux is

$$\Pi_{ik} = \delta_{ik} p + \rho v_i v_k \tag{18.5}$$

The integral of this expression over a closed surface

$$\int \Pi_{ik}\, dS_k = F_i \tag{18.6}$$

is equal to the ith momentum component carried across the surface in unit time. If somewhere within the surface there is a solid body moving in the fluid, the components of vector $\mathbf{F}$ are equal to the force with which the body acts upon the liquid.

At sufficiently large distances from the front of the body the fluid is at rest. At sufficiently large distances behind the body the motion is concentrated mainly within the wake.

Denote the velocity of the body as $\mathbf{v}_0$ and direct the x axis along $\mathbf{v}_0$. Denote the drag by F_x. Assuming the body to be symmetrical with respect to the median plane parallel to the velocity of the body, place the y axis in that plane perpendicular to the velocity $\mathbf{v}_0$. Then F_y is the lift.

Going over to a frame of reference in which the body is at rest, we see that the velocity of unperturbed motion of the fluid is $-\mathbf{v}_0$, and the velocity perturbation caused by the body is denoted $\mathbf{v}'$. Correspondingly, represent the pressure as $p_0 + p'$, where p_0 is constant. The tensor of the momentum flux, Π_{ik}, then takes the form

$$\Pi_{ik} = p_0\delta_{ik} + \rho v_{0i}v_{0k} - \rho v_{0i}v'_k + (p'\delta_{ik} - \rho v'_i v_{0k}) + \rho v'_i v'_k \tag{18.7}$$

The integral of the constant term over the closed surface is zero:

$$\int (p_0\delta_{ik} + \rho v_{0i}v_{0k})\, dS_k = (p_0\delta_{ik} + \rho v_{0i}v_{0k}) \int dS_k = 0$$

This is obvious since $\int dS_k = 0$. The integral of the third term in (18.7) is proportional to the total flux of the fluid across the closed surface:

$$\int \rho v_{0i}v'_k\, dS_k = v_{0i} \int \rho v'_k\, dS_k$$

and, naturally, also vanishes.

At a sufficiently large distance from the body the quadratic perturbation with respect to the velocities v'_i is small in comparison with the linear perturbation. Therefore, the last term on the right in (18.7) can be neglected. Then only the term in parentheses in (18.7) contributes to the total momentum flux:

$$F_i = \int (p'\delta_{ik} - \rho v_{0k}v'_i)\, dS_k \tag{18.8}$$

Let us now take the surface over which the integration is carried out in the form of two planes perpendicular to the x axis: one sufficiently far away in front of the body, the other sufficiently far away behind it.

We shall show that the integral over the front surface is infinitesimal. As the motion outside the wake is conservative, the strong

form of Bernoulli's equation (15.25) holds:

$$p_0+\frac{1}{2}\rho v_0^2=p_0+p'+\frac{\rho}{2}(\mathbf{v}_0-\mathbf{v}')^2 \tag{18.9}$$

Neglecting the term quadratic with respect to the perturbation, and taking into account that $\mathbf{v}_0$ has only one component along the x axis, we obtain

$$p'-\rho v_0 v'_x=0 \tag{18.10}$$

But this is the integrand in (18.8).

A similar reasoning is applicable to the integral over the real surface in a region that does not intersect the wake. There remains only the integral over a cross section of the wake.

The additional pressure p' in the wake is a quantity of the same order of magnitude as outside the wake. We stated this before when examining the pressure in the boundary layer of a body around which a fluid flows. But from (18.10), outside the wake the pressure p', is of the same order of magnitude as the quantity $\rho v_0 v'_x$ in the same domain. And $\rho v_0 v'_x$ outside the wake is, obviously, by the very definition of the wake, less than in the wake. Hence, in Eq. (18.8) it is sufficient to retain only the second term integrated over the cross section of the wake.

As a result the drag F_x is

$$F_x=-\rho v_0\int\int v'_x\,dy\,dz \tag{18.11}$$

The integral divided by v_0 represents the change in the fluid's flow across a cross section of the wake due to the presence of the body. The velocity v'_x is directed in the direction opposite to $\mathbf{v}_0$. Thus, force F_x (the force acting on the fluid), is directed in the same direction in which the body is moving. From Newton's Third Law, there is a force equal in magnitude and opposite in sense acting on the body and retarding its motion. The energy of the body's motion is dissipated mainly in viscous flow within the wake.

A purely conservative flow about the body does not yield a resultant force of resistance. As an example we can investigate such flow around a sphere (Exercise 1, Sec. 15). The absolute value of the velocity is distributed symmetrically with respect to the median plane through the centre of the sphere and perpendicular to the mainstream. Then by Bernoulli's theorem the pressure at corresponding points of the sphere is also symmetric, so that the resultant force (the drag) due to the flow around the sphere is zero. Initially this result, which contradicted experience, seemed incomprehensible. It is called *D'Alembert's paradox*.

Let us also write the expression for lift. According to (18.8) it is equal to

$$F_y = -\int\int \rho v_0 v_y' \, dy \, dz \tag{18.12}$$

This integral is taken over the cross section of the wake. The pressure here was not involved in the initial expression, since the force F_y is tangential to the integration surface while the pressure is perpendicular to it. In other respects (18.12) is derived in approximately the same way as (18.11). On the front surface, where the flow is conservative, we must introduce the velocity potential according to the formula $v_y' = \partial\varphi/\partial y$. Then

$$\int_{-\infty}^{\infty} v_y' \, dy = \varphi(\infty) - \varphi(-\infty) = 0$$

The integral over the rear surface is actually taken only over the cross section of the wake.

Wake Structure. Let us evaluate v_x' at various distances from the body. According to what was said before, the pressure within the wake can be neglected, so that the viscous forces are balanced by inertia forces alone. For v_x' we obtain the approximate equation

$$v_0 \frac{\partial v_x'}{\partial x} = \nu \frac{\partial^2 v_x'}{\partial y^2} \tag{18.13}$$

On the left here we have v_0 instead of $(v_0 - v_x')$. This will be substantiated by our evaluation. Replacing the derivatives by fractions, we arrive at the following evaluation:

$$\frac{v_0 v_x'}{x} \approx \nu \frac{v_x'}{y_0^2}$$

where y_0 is the total width of the wake. This yields

$$y_0 \approx \left(\frac{\nu x}{v_0}\right)^{1/2} \tag{18.14}$$

If the body's lateral dimensions are the same in both directions, integral (18.11) can be written down approximately as

$$F_x \approx \rho v_0 v_x' y_0^2 \tag{18.15}$$

Substituting y_0 from (18.14), we obtain

$$v_x' = \frac{F_x}{\rho \nu x} \tag{18.16}$$

Hence, the velocity perturbation is inversely proportional to the distance from the body. Obviously, at large x we have $v_x' \ll v_0$.

Streamlined Bodies. It will be observed from (18.11) that the narrower the wake the less the drag. For a narrower wake the boundary layer must be made to separate from the body as close as possible to its trailing edge. The boundary layer separates if the downstream pressure increases sharply. The smoother the body's cross section tapers towards the trailing edge, the slower the pressure builds up and the better the boundary layer adheres. Streamlined bodies

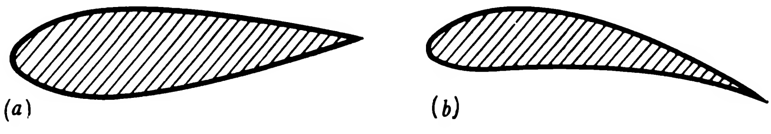


Figure 17

are generally shaped like the profile in Figure 17. They may be bodies of revolution (Figure 17*a*) or bodies elongated considerably perpendicular to the drawing (Figure 17*b*). In flowing around such bodies the streamlines converge smoothly behind the trailing edge and form a narrow wake. The wake cannot be avoided altogether since flow close to the surface obeys the Navier-Stokes equations and possesses a curl **v** which is not zero, and eddies attenuate very slowly.

But if the wake is very thin, the conditions adopted for developing (18.11) are violated. The main portion of the drag is given not by the integral of a member linear with respect to the velocity perturbation over a narrow cross section of the wake but by an integral of quadratic terms over a wide region beyond the wake. This drag will be discussed later in connection with the dynamic lift of streamlined bodies.

The Kutta-Zhukovskii Theorem. Let us now determine the lift force acting on an airfoil. For the sake of simplicity we shall imagine a wing of infinite span and constant longitudinal cross-sectional area. The flow pattern in that case is two-dimensional. The lift per unit length is given by the following integral over the cross section of the wake:

$$F_y = \rho v_0 \int v_0' \, dy \tag{18.17}$$

Unlike the integral in the equation for drag, this integral is of finite value even in the case of an infinitesimally thin wake. Remember that coordinate y is across the wake. Assuming

$$v_y' = \frac{\partial \varphi}{\partial y}$$

where φ is the velocity potential, we see that the integral over the cross section of the wake can be written as

$$F_y = \rho v_0 \int \frac{\partial \varphi}{\partial y} dy = \rho v_0 (\varphi_2 - \varphi_1) \tag{18.18}$$

Here $(\varphi_2 - \varphi_1)$ is the velocity potential gradient between the upper and lower boundaries of the wake.

Assuming the wake to be very thin, we shall treat the difference $(\varphi_2 - \varphi_1)$ as the change in velocity potential in flow around the

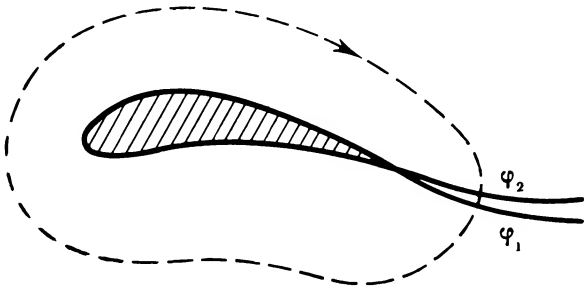


Figure 18

closed dashed line in Figure 18. But this, as was shown in Exercise 2, Section 15, is the velocity circulation Γ in the motion around a closed path. Hence, the lift of the airfoil is

$$F_y = \rho v_0 \Gamma \tag{18.19}$$

(the *Kutta-Zhukovskii theorem*).

For lift to develop, circulation around the airfoil must appear. The direction of circulation should be such that under the airfoil its velocity is subtracted from the velocity of the mainstream, and added to it above the airfoil. Then the resultant velocity under the airfoil will be less than above. From the strong form of Bernoulli's theorem (applicable here, since in a multiply-connected region the circulation may differ from zero when there is no vortex) the pressure under the wing is larger than above it, which generates lift.

Note the difference between formulas (18.11) and (18.12) for drag and lift. The integral (18.12) is taken over a total differential; and remains finite for an infinitesimal wake. The drag equation (18.11) yields values of F_x that decrease as the wake becomes thinner.

Therefore, as indicated before, for well-streamlined bodies we must go over to the contribution to the drag made by the velocity field outside the wake.

We shall define this drag without resorting to formulas. It can be expressed linearly in terms of the derivatives of the circulation Γ

taken along the airfoil, that is, with respect to z. Hence, if an airfoil is of infinite span, so that Γ does not depend on z, the drag referred to unit length tends to zero. Then the drag must be calculated partially according to Eq. (18.11) and partially according to viscous friction in the boundary layer at the surface of the airfoil. One way or the other, it turns out to be many times less than the lift. Obviously this is a necessary condition for flight.

On the Calculation of Circulation. To calculate lift it is necessary to know the velocity circulation with respect to the airfoil. Generally speaking, the problem does not have a unique solution for the case of arbitrary flow. However, we may require that the flow meet smoothly behind the trailing edge without carrying off additional vortices. Such vortices could be expected to form in the meeting of flows of different velocities perpendicular to the direction of flow (a discontinuity in the tangential component of any vector on a surface is equivalent to a surface vortex). The condition for the absence of vortices behind the trailing edge of an airfoil was set by S. A. Chaplygin and N. E. Zhukovskii.

In this case the problem of flow around an airfoil is solved using functions of a complex variable described in Section 15.

Suppose a complex velocity potential is given by a function $w(z)$ (the airfoil is assumed to have infinite span). Determination of the function $w(z)$ for a specific cross section is an extremely difficult task. Let us therefore consider it in general form.

Assuming the airfoil to be at rest and the air to be flowing past it, the velocity of the air must be taken as constant at infinite distance from the airfoil. Like the complex potential w, we can develop the complex velocity, dw/dz. Then, regardless of the choice of dz, the velocity components are equal to the real and imaginary parts of the derivative dw/dz, provided the Cauchy-Riemann equations are satisfied.

Let us find the dependence of dw/dz on z. For the function to take a finite value at $|z| \to \infty$, it must be sought in the form of a series expansion in the reciprocal powers of z:[6]

$$\frac{dw}{dz} = A + \frac{B}{z} + \frac{C}{z^2} + \ldots$$

Integrating, we obtain

$$w = \psi + i\varphi = Az + B \ln z - \frac{C}{z} + \ldots \qquad (18.20)$$

[6] These must be integral powers of z, since fractional powers lead to ambiguity (for example, $\sqrt{z}$ has two signs), while velocity must be a unique function of coordinates.

The first term in this formula corresponds to the complex potential of steady flow. The second term gives the perturbation caused by the airfoil, which does not tend to zero at infinite values of $|z|$. In Exercise 2, Section 15, it was shown that the potential $\psi + i\varphi = B \ln z$ can correspond to both the source (if B is a real number) and a vortex (if B is purely imaginary). But an airfoil cannot be a source, hence B is a purely imaginary quantity. It must be equal to $\Gamma/(2\pi i)$, in which case the real velocity potential should be ψ.

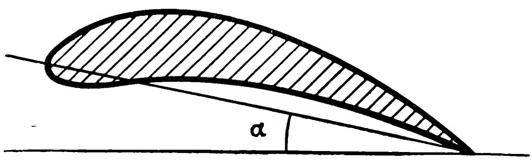


Figure 19

Hence, circulation is determined by the first term of the expansion $w(z)$ for the given airfoil cross section. The Chaplygin-Zhukovskii condition is necessary for $w(z)$ to be uniquely defined over the given cross section.

There is always a position of an airfoil in a flow at which the lift becomes zero. For example, if an airfoil has a plane of symmetry with respect to the upper and lower sides, and the velocity of motion of the airfoil lies in that plane (Fig. 17*a*), then obviously there is no lift.

In the case of a nonsymmetrical airfoil there also exists a plane motion in which lift is zero. The angle between the velocity of motion of an airfoil and that plane is called the *angle of attack*, or *angle of incidence*, α (Figure 19). At small angles of attack the lift is proportional to α. At large values of α smooth flow around the cross section becomes impossible. The Zhukovskii-Chaplygin principle (or condition) cannot be satisfied at the trailing edge, and a large drag develops.

There exists an optimum choice of α, since at too small values of α the lift may prove insufficient.

Incorrect choice of the angle of attack was the cause of numerous failures in the early days of aviation.

EXERCISES

1. Show that the term $v'_y\,(\partial v'_x/\partial y)$, neglected in evaluating the wake width, is of the same order as $v'_x\,(\partial v'_x/\partial x)$ with respect to the principal term $v_0\,(\partial v'_x/\partial x)$.

Solution. From (18.16) $v'_x = F_x/(\rho\nu x)$. But $\frac{\partial v'_x}{\partial x} + \frac{\partial v'_y}{\partial y} = 0$. Whence

$$\frac{\partial v'_y}{\partial y} = \frac{F_x}{\rho\nu x^2}, \quad \text{or} \quad v'_y \approx \frac{F_x y_0}{\rho\nu x^2}$$

Since

$$\frac{\partial v'_x}{\partial y} \approx \frac{v'_x}{y_0}$$

we find that

$$v'_y \frac{\partial v'_x}{\partial y} \approx \frac{F_x y_0}{\rho\nu x^2} \frac{v'_x}{y_0} = \frac{F_x^2}{(\rho\nu)^2 x^3}$$

The ratio of this discarded term to the principal term is

$$\frac{F_x}{v_0 \rho\nu x} \approx \frac{v'_x}{v_0}$$

From Eq. (18.16) one can see that at a sufficient distance from the body v'_x is always less than v_0.

2. Plane flow occurs around an infinitely long cylinder of radius a, perpendicular to the axis. The velocity of the flow at infinity is v_0, the circulation Γ around the cylinder is given. Determine the lift, assuming the fluid to be ideal and the flow irrotational.

Solution. Since the flow is irrotational, there exists a velocity potential satisfying the Laplace equation $\nabla^2\varphi = 0$. Since the equation is linear, the superposition principle holds [Sec. 15]. Laminar flow corresponds to the velocity potential

$$\varphi = v_0 \left(r + \frac{a^2}{r}\right) \cos\theta$$

which assures that the radial component vanishes at the cylinder's surface (the angle θ is measured in the direction of the velocity). Adding the circulation component of the velocity to v_θ, we obtain the velocity field.

We find the pressure at the surface of the cylinder from the strong form of Bernoulli's theorem:

$$p = -\rho \frac{v^2}{2}$$
$$= -\frac{\rho}{2} \left\{ v_0^2 \left(1 - \frac{a^2}{r^2}\right) \cos^2\theta + \left[v_0 \sin\theta \left(1 + \frac{a^2}{r^2}\right) - \frac{\Gamma}{2\pi r} \right]^2 \right\}$$

To obtain the lift from this we must calculate the integral

$$F_y = -\int_0^{2\pi} \rho \frac{v^2}{2} a \sin\theta \, d\theta = \rho\Gamma v_0$$

The other pressure components are orthogonal to $\sin\theta$ within $[0, 2\pi]$. The formula obtained here[7] agrees with the general theorem of Kutta-Zhukovskii

[7] It was derived by Rayleigh long before the general formula.

even though it refers to other flow conditions than those for which it was proved in the text.

The drag is equal to

$$F_x = \int_0^{2\pi} \rho \frac{v^2}{2} a \cos\theta \, d\theta = 0$$

as it should be in irrotational flow of an ideal liquid around a body. This result was called D'Alembert's paradox in the text.

19

SUPERFLUIDITY

Quantum Liquid. At atmospheric pressure helium remains liquid down to absolute zero. Qualitatively this can be explained in the following way. As was shown in [Sec. 28], bound states of a particle (states in which it moves finitely) do not occur for all attractive potentials but only when the condition

$$\frac{mUa^2}{\hbar^2} \gtrsim 1 \tag{19.1}$$

where U is the effective depth, and a is the radius of the potential well, is satisfied.

The persistence of the liquid state of helium is an indication that this condition is not satisfied owing to the low mass of the atom and the small depth of the well. Hydrogen, as is known, goes over to the solid state of a molecular crystal. Therefore, small atomic mass is not a sufficient condition for a substance to remain liquid down to absolute zero. Apparently, helium atoms, with their closed (spherically symmetrical) electron shells, interact weaker than hydrogen molecules, and condition (19.1) is satisfied for hydrogen, but not for helium.

In a hydrogen molecule there is a high admixture of one-electron atomic states, and therefore two hydrogen molecules interact stronger than two helium molecules, at least in a liquid medium, that is, in the condensed phase. At a pressure of only 25 atm (relatively small for a liquid) helium at absolute zero goes over to the solid state. This shows that for helium, too, condition (19.1) is almost satisfied.

There are two stable isotopes of helium: of atomic weight 4 and atomic weight 3. In natural mixtures the proportion of the lighter

isotope is about one-millionth. The mixture can be separated to obtain pure He^3. Obviously, if condition (19.1) is not satisfied for the heavy isotope, it is even less satisfied for the lighter one. Indeed, both remain liquid down to absolute zero. They are called quantum liquids, because they owe the persistence of their liquid state down to absolute zero to the quantum properties of atomic motion.

Nevertheless, in liquid state the two isotopes behave in an entirely different manner. As was shown in Section 5, liquid He^4 at temperature 2.2 K becomes a superfluid, a state in which it is capable of passing through the thinnest capillaries, displaying no apparent viscosity. Helium atoms of atomic weight 4 possess no nuclear spin and accordingly obey Bose statistics. Therefore below the transition temperature some of the atoms are in the zero-energy state. This should be understood in the sense of the quantum superposition principle: the zero-energy state is added to the state of each atom and is involved in the probability amplitude of each state. He^3 does not become superfluid, which can be correlated with the fact that the spin of an He^3 nucleus is equal to 1/2. Accordingly, He^3 is subject to Fermi statistics. If both isotopes were gases, the question of the difference between them would be easily resolved: a Fermi gas does not tend to accumulate in one state, since this would contradict Pauli's exclusion principle.

However, at present there is no model microscopic theory of a quantum liquid. That is why the difference in the behaviour of the two isotopes offers no more than a strong indication that there is a connection between the type of atomic statistics and superfluidity. But there is no complete proof of this assumption.

There is, however, a phenomenological theory of superfluidity, enunciated by L. D. Landau, which is analogous to the conventional hydrodynamics of an ideal fluid but which gives a complete description of the superfluid properties of macroscopic motion. Landau did not consider atomic statistics.

In this theory, called the two-fluid model, He^4 in its superfluid state is represented as a liquid whose motion is defined not by one velocity but by two, *superfluid* and *normal*. Later the meaning of this will be explained. The equations and conclusions of the two-fluid theory are quite unequivocal and agree beautifully with experimental data.

The Spectrum of Liquid He^4. The phenomenological theory does not derive the properties of the liquid from the microscopic properties of individual atoms but postulates the laws of motion of a quantum liquid as a whole, on the basis of experience. The liquid is treated, as is generally the case in hydrodynamics, as a homogeneous medium. If its motion is quantized, it possesses a certain assembly of quantum states.

Since the medium is homogeneous, each state must be characterized by its own value of momentum (in accordance with the general theorems of mechanics). At temperatures close to absolute zero, when there are few excited states, any one of them may be considered to be independent of the others.

The zero-energy level for such excitations is the energy of a Bose-Einstein condensate, with which the superfluid portion of the liquid is associated, the normal portion being associated with excitation. We repeat that it is impossible to regard certain atoms as belonging to the superfluid portion and others as belonging to the normal portion. The excitations are collective and describe the motion of the fluids as a whole.

The picture of collective excitations was already employed in Section 4 in investigating phonons in a crystal lattice. All the atoms of a lattice take part in the propagation of a wave through it, so that the phonon is an example of joint excitation. Since a lattice has a discrete structure, a phonon possesses not a momentum varying from 0 to ∞ but a wave number with an upper limit. The energy $\hbar\omega$ of a phonon depends on the wave number. This dependence is called the *phonon spectrum*, or the *collective-lattice-excitation* spectrum.

In the same way, collective excitations in liquid helium are characterized by a spectrum: a function stating the dependence of the energy on momentum.

A spectrum of elementary excitations can be found experimentally by observing the scattering of very slow neutrons in liquid helium on elementary excitations. For this it is sufficient to measure the energy of a neutron together with its deflection angle in the same scattering act. Then from the energy and momentum conservation laws the same quantities are determined for a "particle" of the scatterer [Sec. 5]. In other words, the spectrum of excitations on which the neutrons are scattered is restored. That was how the excitation spectrum in liquid He^4, predicted by Landau on the basis of the macroscopic properties of superfluid helium, was confirmed.

Like the phonon of acoustic lattice vibrations, at low momentum the excitation energy in liquid helium is a linear function of the momentum:

$$\varepsilon = cp \tag{19.2}$$

Here, c is a quantity analogous to the velocity of sound. It is natural to assume that the lowest excitation of liquid helium is acoustic. Such an excitation, as in a lattice, is conventionally called a *phonon*. At higher momenta $\varepsilon(p)$ was found to be not a monotonic function of p: it attains a maximum, then decreases and passes through a minimum, after which it again increases. When the momentum is of the order of h/a, where a is the dimension of the atom, the spectral curve terminates. At such short wavelengths a liquid can no longer be treated as a continuous medium.

Capillary Flow. The phenomenon of superfluidity, which was discovered by P. L. Kapitza, consists in that liquid helium flows practically instantaneously through a capillary so thin that flow through it at a point above the transition point would take an extremely long time. Landau explained it on the basis of the form of the spectrum (19.2), as follows.

Let liquid helium at absolute zero be flowing through a capillary with a velocity $\mathbf{v}$. The phenomenon of viscosity consists in that, owing to friction along the wall, the steady motion of a fluid becomes unsteady thermal motion. In terms of quantum states this means that phonons are emitted in the capillary so that a portion of the superfluid liquid goes over to the normal state. Let us consider the conditions in which such a transition is compatible with the conservation laws.

Take a frame of reference in which the helium is at rest and the capillary is moving with the velocity $-\mathbf{v}$. In this frame the energy of the flowing helium prior to a phonon emission is zero, and after the emission of a phonon with momentum $\mathbf{p}$ it is equal to cp.

Now we go back to the old frame of reference. The momentum of helium with respect to this frame is $\mathbf{p}' = \mathbf{p} + m\mathbf{v}$, where m is the mass of the flowing helium. Let us also write the formula for the energy in the moving frame. Since (19.2) has the unusual form of the dependence of energy on momentum, it is simpler to proceed from the Lorentz transformation [14.18] and go over to a nonrelativistic approximation. This is by no means an obligatory method, but it does not involve an error. Representing energy as $mc_0^2 + E$, we write

$$mc_0^2 + E = \frac{mc_0^2 + \varepsilon + \mathbf{pv}}{(1 - v^2/c_0^2)^{1/2}} \tag{19.3}$$

Here c_0 is the velocity of light.

Expanding the denominator into a series and retaining the first term, we obtain, after subtracting mc_0^2 from both sides of the equation, the following formula:

$$E = \varepsilon + \mathbf{pv} + \frac{mv^2}{2} \tag{19.4}$$

Here, E denotes the kinetic energy of the flowing helium after phonon emission, that is, after dissipation. It is less than the initial kinetic energy of flow, $mv^2/2$. Therefore

$$\varepsilon + \mathbf{pv} < 0 \tag{19.5}$$

or, from (19.2),

$$cp + \mathbf{pv} < 0 \tag{19.6}$$

The quantity cp is essentially positive. Hence, if $v < c$, the inequality cannot be satisfied—phonon emission is prohibited by the

conservation laws. The helium flows through the capillary without friction. If the helium is not at absolute zero, all we have to do is investigate the emission of one more phonon in excess of those present, which does not significantly alter the situation. The superfluid component is not subject to viscous forces. The normal component, naturally, displays viscosity, owing to the scattering of elementary excitations on the capillary walls.

It will be shown in Exercise 1 that in the superfluid component the formation of excitations close to the minimum of the spectral curve $\varepsilon(p)$ is impossible if the velocity of flow is below a certain value.

The Condition of Mechanical Equilibrium of the Superfluid Component. Before Kapitza's discovery of the superfluidity of liquid helium the viscosity of helium was measured by observing the damping of the torsional vibrations of a disk immersed in the liquid. Since there was no separation between the superfluid and normal components, all that was measured in such experiments was the viscosity of the normal component.

Only the superfluid component, possessing no viscosity, passes through a very thin capillary. As this component carries no heat (a Bose-Einstein condensate possesses zero energy), the liquid helium remaining in the vessel heats up—the same energy is distributed over a smaller mass.

Consider now two vessels containing liquid helium below the transition point and connected by a very thin capillary. The temperatures in the two vessels are different. Heat is transmitted through the capillary very slowly. Therefore, mechanical equilibrium, maintained by the free mixing of the superfluid component in the capillary, must set in in the first place.

The equilibrium condition is, as always, that work cannot be done in the flow of this component. Since no heat transfer occurs, the work is equal to the change in energy. If the energy of the liquid helium in one vessel is E_1, and E_2 in the other, then the work is equal to

$$A = \Delta (E_1 + E_2) = 0 \tag{19.7}$$

But the change in energy is due solely to the flow of the superfluid component, that is, without any change in entropy. If ΔN_1 particles of helium have flowed out of the first vessel, then

$$\Delta E_1 = \Delta N_1 \left(\frac{\partial E_1}{\partial N} \right)_S = \mu_1 \Delta N_1 \tag{19.8}$$

since $dE = \theta\, dS - p\, dN + \mu\, dN$ (see Eq. (8.52)). A similar relationship can be written down for ΔE_2. But $\Delta N_1 = -\Delta N_2$. Therefore

$$\Delta N\, (\mu_1 - \mu_2) = 0 \tag{19.9}$$

Finally, the condition of mechanical equilibrium consists in that the chemical potentials of the helium in both vessels become equal:

$$\mu(p_1, \theta_1) = \mu(p_2, \theta_2) \tag{19.10}$$

Accordingly, at different temperatures the helium assumes different heights in the vessels, depending upon the pressure. Let us note that for the normal component the equilibrium condition consists in the equality of pressures.

The Linearized Hydrodynamic Equations of Liquid Helium. The exact hydrodynamic equations of liquid helium are very complex in form. Besides, they include one theoretically indeterminate function: the density ratio of the superfluid and normal components as a function of their relative velocity. We shall restrict ourselves to the consideration of only acoustic waves in liquid helium. The linearized form of the equations (Sec. 16) is sufficient for this.

We represent the density of the liquid flow in the form

$$\mathbf{j} = \rho_n \mathbf{v}_n + \rho_s \mathbf{v}_s, \tag{19.11}$$

where ρ_n and $\mathbf{v}_n$ are the density and velocity of the normal component, and ρ_s and $\mathbf{v}_s$ are the density and velocity of the superfluid component. Assuming that $\mathbf{v}_n$ and $\mathbf{v}_s$ are small, we find that it is sufficient to consider ρ_n and ρ_s to depend only on the temperature, as in a liquid at rest. Obviously $\rho_n + \rho_s = \rho$, where ρ is the total density.

Comparing (19.11) with the expression for the momentum density involved in (15.15), we find that they are identical, that is, $\mathbf{j}$ can be taken for the momentum density. But the expression for the density of the momentum flux in (15.15) involves a nonlinear term, $\rho v_i v_k$, which should be discarded in a linearized equation. Thus, we obtain

$$\frac{\partial \mathbf{j}}{\partial t} = -\operatorname{grad} p \tag{19.12}$$

We must next take into account that entropy is transferred only by the normal component of velocity. Denoting the entropy per unit mass by the letter S, the entropy per unit volume is ρS, and the density of the entropy flux is $\rho \mathbf{v} S$. Here, by the definition of S, it is ρ and not ρ_n that is involved, since the entropy is taken per unit of the total mass. Neglecting dissipative processes, we find that the entropy satisfies a conservation law similar to the mass and charge conservation laws:

$$\frac{\partial}{\partial t}(\rho S) + \rho S \operatorname{div} \mathbf{v}_n = 0 \tag{19.13}$$

Now we must write the dynamic equation for the velocity of the superfluid component. Mechanical equilibrium for this component is

achieved at μ = constant. Hence, at small velocities the acceleration of the superfluid component is proportional to $-\mathrm{grad}\,\mu$. Here μ is the chemical potential referred to unit mass. Let us prove that the proportionality is unity. Indeed, if we write

$$\frac{\partial \mathbf{v}_s}{\partial t} = -\mathrm{grad}\,\mu \tag{19.14}$$

and multiply both sides of the equation by $\mathbf{v}_s$, on the left we obtain $(\partial/\partial t)v_s^2/2$. The quantity $-\mathrm{grad}\,\mu$ is the energy gradient referred at constant entropy to unit mass, that is, force per unit mass. The product of this force and the velocity $\mathbf{v}_s$ equals the kinetic energy increment of unit mass in unit time, as it should be.

Finally, the fourth equation has the same meaning as (15.6), that is, it expresses the law of conservation of mass:

$$\frac{\partial \rho}{\partial t} + \mathrm{div}\,\mathbf{j} = 0 \tag{19.15}$$

(the continuity equation).

Equations (19.12)-(19.15) constitute a complete set: they contain four unknown quantities, $\mathbf{j}$, $\mathbf{v}_s$, p and μ. The other quantities, ρ, S, and $\mathbf{v}_n$, are expressed directly in terms of the former quantities: entropy and density according to thermodynamic formulas, and $\mathbf{v}_n$ from Eq. (19.11), provided the ratio ρ_s/ρ_n in helium at rest is known.

Second Sound. We shall now obtain the equations describing the propagation of an acoustic wave in liquid helium. First, we take the time derivative of (19.15) and replace $\partial \mathbf{j}/\partial t$ with the help of (19.12) to obtain

$$\frac{\partial^2 \rho}{\partial t^2} = \nabla^2 p \tag{19.16}$$

Now we find the time derivative of S:

$$\begin{aligned}\frac{\partial S}{\partial t} &= \frac{1}{\rho}\frac{\partial}{\partial t}\rho S - \frac{S}{\rho}\frac{\partial \rho}{\partial t}\\ &= -\frac{1}{\rho}\rho S\,\mathrm{div}\,\mathbf{v}_n + \frac{S}{\rho}\mathrm{div}\,(\rho_n\mathbf{v}_n + \rho_s\mathbf{v}_s)\end{aligned}$$

Here, ρ_n and ρ_s need not be differentiated (see (19.13)); therefore

$$\frac{\partial S}{\partial t} = \frac{S\rho_s}{\rho}\,\mathrm{div}\,(\mathbf{v}_s - \mathbf{v}_n) \tag{19.17}$$

Making use of the fact that $\mu = G/N$, we obtain from the thermodynamic identity (8.45) for G:

$$\mathrm{grad}\,\mu = -S\,\mathrm{grad}\,\theta + \frac{1}{\rho}\mathrm{grad}\,p \tag{19.18}$$

Next we must substitute grad p and grad μ from the equations of motion (19.12) and (19.14), and we obtain

$$-\frac{\partial \mathbf{v}_s}{\partial t} = -S \operatorname{grad} \theta - \frac{1}{\rho}\left(\rho_n \frac{\partial \mathbf{v}_n}{\partial t} + \rho_s \frac{\partial \mathbf{v}_s}{\partial t}\right)$$

Finally, making use of the fact that $\rho_n + \rho_s = \rho$, we obtain

$$\frac{\rho_n}{\rho} \frac{\partial}{\partial t} (\mathbf{v}_n - \mathbf{v}_s) = -S \operatorname{grad} \theta \tag{19.19}$$

Now, excluding $(\mathbf{v}_n - \mathbf{v}_s)$ from (19.17) and (19.19), we arrive at the equation

$$\frac{\partial^2 S}{\partial t^2} = \frac{S^2 \rho_s}{\rho_n} \nabla^2 \theta \tag{19.20}$$

Equations (19.16) and (19.20) contain only thermodynamic variables, two of which are, as always, independent.

We shall now show that at low temperatures these equations describe different, nonrelated wave processes. The compressibility of a condensed substance at low temperature is due to resilient forces among molecules. That is why the density is dependent mainly upon the pressure that compresses the liquid. Consequently, $(\partial^2 \rho/\partial t^2)$ should be replaced by $(\partial \rho/\partial p)(\partial^2 p/\partial t^2)$, the differentiation of density being carried out at constant entropy or temperature, which in the given conditions does not matter. The effect of thermal excitations on compressibility at low temperatures is insignificant since there are very few excitations. But this yields the normal equation (16.6) for the propagation of acoustic vibrations in a liquid.

The entropy of liquid helium per unit mass, S at low temperatures depends mainly on the temperature (see Exercise 2). As in the case of a solid body, the phonon part of the entropy is proportional to the cube of the temperature. Besides, at a temperature of 1 K the contribution of excitations whose energy is close to the minimum on the curve $\varepsilon(p)$ still exists.

Substituting $(\partial S/\partial \theta)_\rho (\partial^2 \theta/\partial t^2)$ for $(\partial^2 S/\partial t^2)$, we arrive at a wave equation of the form

$$\frac{\partial^2 \theta}{\partial t^2} = \frac{S^2 \rho_s}{\rho_n} \frac{\theta}{C_V} \nabla^2 \theta \tag{19.21}$$

This type of acoustic excitations is called *second sound*. It was predicted by Landau on the basis of his equations (19.12)-(19.15).

Let us investigate these vibrations in greater detail. They take place at constant volume or pressure, which in the present case is of no consequence. From Eq. (19.12) we can see that constant pressure corresponds to $\mathbf{j} = 0$, that is, to the condition $\rho_n \mathbf{v}_n + \rho_s \mathbf{v}_s = 0$. The liquid as a whole remains at rest while the superfluid component oscillates with respect to the normal component. At the transition

point, where ρ_s becomes zero, the second sound also disappears, which is directly apparent from (19.21).

Where the concentration of thermal excitations increases, the temperature, naturally, increases together with ρ_n. Therefore, second sound corresponds to a wave of temperature vibrations propagating through the liquid helium.

Accordingly, E. M. Lifshitz suggested exciting such vibrations with the help of an electric heating coil through which alternate current is passed. V. P. Peshkov used the method to observe second sound.

Quantized Vortices. We shall now examine the wave function of the motion of a quantum liquid as a whole. A wave function of this kind can always be isolated from the general wave function in the form of a factor. If the velocity of macroscopic flow is **V**, the wave function can be written as

$$\psi = e^{iM\mathbf{Vr}/\hbar} \tag{19.22}$$

where M is the mass of the liquid. We can write ψ differently in the following form:

$$\psi = \exp\left(\frac{i}{\hbar}\sum m\mathbf{Vr}\right) = \prod e^{im\mathbf{Vr}/\hbar} \tag{19.23}$$

Here, Π denotes the product over all the atoms. Now, let the velocity not be strictly constant over the volume, varying slowly from point to point at atomic distances. Then (19.23) can be replaced by an approximate expression of the form

$$\psi = \prod \exp\left(i\frac{m}{\hbar}\int \mathbf{V}\,d\mathbf{r}\right) \tag{19.24}$$

where Π again denotes the product over individual atoms.

Let us apply this formula to the circular motion of the liquid, when the velocity integrated along a closed path is not zero. Since the wave function is single valued, the integral in (19.24) can vary along the closed path only by an integer that is a multiple of $2\pi\hbar/m$. In other words, the circulation along the closed path can assume only the following values:

$$\Gamma = \int \mathbf{V}\,d\mathbf{r} = \frac{2\pi\hbar}{m}n \tag{19.25}$$

with n an integer.

But the motion of a superfluid liquid is predominantly potential motion, since this corresponds to the smallest excitations (for example, acoustic disturbances are potential). In Exercise 2, Section 15, it was shown that potential motion of a fluid can yield a nonzero circulation only if there is a vortex line in it.

It is most interesting that from (19.25) a vortex line yielding a circulation Γ is characterized by a quantity of macroscopic order. Indeed, since $m = 6.4 \times 10^{-24}$ g, $\Gamma \sim 10^{-3}\ \text{cm}^2\text{s}^{-1}$. At a distance 10^{-6} cm from the vortex's axis, which is great in comparison with atomic dimensions, this corresponds to a circulation velocity of $10^2\ \text{cm-s}^{-1}$, which can be observed directly by the scattering of slow (thermal) neutrons.

The existence of quantized vortices explains how the rotation of a vessel is transmitted to the liquid helium it contains.

From Eq. (19.6) superfluid motion of a liquid through a capillary is possible up to velocities equal to c. Actually it ceases at much lower velocities, evidently because of formation of quantized vortices, predicted independently by Lars Onsager and Richard P. Feynman.

EXERCISES

1. Prove that at a flow velocity below a certain value excitation states cannot form in a superfluid liquid close to the minimum of the potential curve given by the formula

$$\varepsilon(p) = \Delta + \frac{(p-p_0)^2}{2M_0}$$

2. Determine the expression for the energy of liquid helium at a low temperature, taking also into account the states considered in Exercise 1.

Solution. The excitations behave like a Bose gas with $\mu = 0$, for which

$$E = \int \frac{\varepsilon(p)}{e^{\varepsilon/\theta}-1} \frac{V\,d\mathbf{p}}{(2\pi h)^3} = -\theta^2 \frac{\partial}{\partial\theta} \int \ln\left(1-e^{-\varepsilon/\theta}\right) \frac{V\,d\mathbf{p}}{(2\pi h)^3}$$

Two regions make contributions to the integral: close to zero and close to $p = p_0$. Extending both integrals over infinite limits, which is valid at low temperatures, we obtain

$$E = -\theta^2 \frac{\partial}{\partial\theta} \left\{ \int \ln\left(1-e^{-cp/\theta}\right) \frac{V\,d\mathbf{p}}{(2\pi h)^3} + \int \ln\left[1-\exp\left(-\frac{\Delta}{\theta}-\frac{(p-p_0)^2}{2M_0\theta}\right)\right] \frac{V\,d\mathbf{p}}{(2\pi h)^3} \right\}$$

Since $\Delta \gg \theta$, in the second integral we can replace the logarithm with the first term of its expansion into a series. Then

$$E = -\theta^2 \frac{\partial}{\partial\theta} \left[\frac{V\theta^3}{2\pi^2 h^3 c^3} \int_0^\infty \ln\left(1-e^{-x}\right) dx + \frac{V\Delta p_0^2 (2\pi M_0\theta)^{1/2}}{2\pi^2 h^3} e^{-\Delta/\theta} \right]$$

Taking into account that

$$\int_0^\infty \ln(1-e^{-x})\,dx = -\sum_{n=1}^\infty \frac{1}{n^2} = -\frac{\pi^2}{6}$$

(see Appendix to Part I), we finally obtain

$$E = \frac{V\theta^4}{4h^3c^3} + \frac{V\Delta p_0^2 (2\pi M_0\theta)^{1/2}}{2\pi^2 h^3} e^{-\Delta/\theta}$$

where[8] $\Delta/k_B = 8.6$ K (or $\Delta = 1.2 \times 10^{-15}$ erg), $p_0 = 2.0 \times 10^{-19}$ g-cm-s^{-1}, $M_0 = 0.16 m_{He^4} = 0.105 \times 10^{-23}$ g. At a temperature of 1 K

$$\frac{E}{V} = (0.74 + 0.12) \times 10^4 \text{ erg-cm}^{-3}$$

The contribution of excitations with momentum in the neighbourhood of p_0 is 14%. To calculate the specific heat the obtained expression must be differentiated with respect to θ. In this, the first term will be multiplied by 4, and the second by Δ/θ, or by 8.6. Hence, the fraction of excitations with $p \sim p_0$ in the specific heat is around 29%.

20

ONE-DIMENSIONAL STEADY FLOW OF A COMPRESSIBLE GAS

Thermodynamic Quantities. This section commences an examination of the flow of a compressible gas. As was shown in Section 16, fluid compressibility is an important factor at flow velocities approaching the speed of sound or exceeding it. It is therefore useful to have expressions for the thermodynamic quantities describing the state of a gas in terms of the speed of sound propagating through it. Usually thermodynamic formulas for quantities defining the state of a gas are extremely cumbersome and special tables must be used[9].

But very frequently gases are considered whose specific heat ratios are constant, as for example air at temperatures ranging from 200 to 1500 K (approximately). The vibrational degrees of freedom of

---

[8] For the values of the constants see D. Henshaw and A. Woods, "Modes of atomic motions in liquid helium II", *Phys. Rev.*, **121**, 1266 (1961).

[9] See N. M. Kuznetsov, *Termodinamicheskie funktsii i udarnye adiabaty vozdukha pri vysokikh temperaturakh* (Thermodynamic Functions and Hugoniot Adiabats of Air at High Temperatures), Mashinostroenie Press, Moscow, 1965.

air are almost unexcited even at $c_p/c_V = 7/5$. In monatomic gases the constancy of specific heat is disturbed only by the excitation of electronic degrees of freedom, and $c_p/c_V = 5/3$ in an even broader temperature range.

The velocity of sound, c, is given by the formula (16.7):

$$c^2 = \left(\frac{\partial p}{\partial \rho}\right)_S \tag{20.1}$$

As was shown in Exercise 1, Section 8, the isentropic derivative $(\partial p/\partial \rho)_S$ is connected with the isothermal derivative by the relationship[10]

$$\left(\frac{\partial p}{\partial \rho}\right)_S = \frac{c_p}{c_V}\left(\frac{\partial p}{\partial \rho}\right)_\theta \equiv \frac{c_p}{c_V}\left(\frac{\partial p}{\partial \rho}\right)_T \tag{20.2}$$

The derivative $(\partial p/\partial \rho)_T$ is determined from the ideal gas law (2.24), in which the gas constant must be taken not per mole but per gram of substance, that is, not R but R/M, where M is the molecular weight of the substance. Then

$$p = \frac{RT}{M}\rho \tag{20.3}$$

Note that here

$$\rho \equiv \frac{1}{V} \tag{20.4}$$

where V is the volume of one gram, or specific volume, of the gas. Therefore

$$\left(\frac{\partial p}{\partial \rho}\right)_T = \frac{RT}{M} \tag{20.5}$$

or, denoting $c_p/c_V \equiv \gamma$, we obtain

$$c^2 = \frac{c_p}{c_V}\,\frac{RT}{M} = \gamma\,\frac{RT}{M} \tag{20.6}$$

A gas with a constant specific heat ratio obeys the equipartition principle, and the energy and enthalpy of one gram of such gas are given, up to a constant term, by the formulas

$$E = \frac{c_V T}{M}, \qquad H = \frac{c_p T}{M} \tag{20.7}$$

Taking into account that $c_p - c_V = R$ (for a gas) and using (20.6), we obtain the expressions for the specific energy and specific en-

---

[10] In this section and subsequently the temperature T is in kelvins.

thalpy in terms of the velocity of sound:

$$E = \frac{c^2}{\gamma(\gamma-1)} \tag{20.8}$$

$$H = \frac{c^2}{\gamma-1} \tag{20.9}$$

The Significance of the Velocity of Sound in the Dynamics of Compressible Gas. The velocity of sound c in Eq. (20.1) is defined in a frame of reference in which the gas is at rest at a given point. Hence, according to Eqs. (20.8) and (20.9), it characterizes the internal, or thermodynamic, state of the gas. The velocity of sound in a gas with respect to a fixed reference frame can be obtained by adding the velocity of the gas proper to the velocity of sound in it.

In the dynamics of a compressible gas, or *gas dynamics*, as it is conventionally called, the velocity of sound is of the same significance as the velocity of light in electrodynamics. Disturbances, or "signals", are transmitted from one part of the gas to another at the speed of sound. Unlike electrodynamics, where the concept of faster-than-light speed is found only in the exceptional case of the Cerenkov effect (Sec. 39), in gas dynamics the velocity of matter is very often greater than that of sound signals. Bodies travelling through a medium (bullets, artillery shells, missiles, aircraft), and a substance itself, usually a gas in a stationary pipe, can move at supersonic velocities.

If the gas encounters a small obstacle, the disturbance is transmitted relative to the gas with the speed of sound. But when the velocity of flow is greater than the velocity of sound the disturbance cannot be transmitted upstream. The gas impinging on the obstacle has no "knowledge" of what lies in its way. Contrariwise, in flow around a body at subsonic speeds disturbances propagate infinitely upstream.

The Limiting Velocity. In steady isentropic flow the weak form of Bernoulli's theorem (15.17) holds:

$$\frac{v^2}{2} + H = \text{constant} \tag{20.10}$$

If a gas is flowing from a vessel where it is at rest and possesses enthalpy H_0, going over to a state with enthalpy H, from (20.10) we have

$$\frac{v^2}{2} + H = H_0 \tag{20.11}$$

whence the velocity in the new state is

$$v = [2(H_0 - H)]^{1/2} \tag{20.12}$$

Suppose a gas is flowing into a vacuum. In [Sec. 8] we treated this as an irreversible process. But irreversibility appears only if the outflowing gas comes to rest, when its kinetic energy of ordered motion transforms into its internal energy E. As long as the gas is in motion its entropy does not change. We express it via the pressure and temperature of the gas with the help of (9.28):

$$S = c_p \ln T - R \ln p \tag{20.13}$$

From this, taking into account that

$$\frac{c_p}{R} = \frac{c_p}{c_p - c_V} = \frac{\gamma}{\gamma - 1}$$

we obtain the equation for an isentropic process:

$$\frac{T^{\gamma/(\gamma-1)}}{p} = \text{constant} \tag{20.14}$$

Thus, in isentropic expansion into vacuum a gas cools to zero temperature (this property is used to liquefy real gases). At $p = 0$ the enthalpy is, from (20.7), also zero. It follows that the greatest velocity of steady outflow corresponds to a transition into vacuum and is equal to

$$v_0 = (2H_0)^{1/2} \tag{20.15}$$

or, with the help of (20.9),

$$v_0 = \left(\frac{2}{\gamma - 1}\right)^{1/2} c_0 \tag{20.16}$$

This formula shows the advantage of expressing enthalpy directly in terms of the velocity of sound.

For air, $v_0 = c_0\sqrt{5}$.

The Critical Velocity. Since the velocity of sound in a compressible gas changes from point to point, the determination of subsonic and supersonic flow is localized. At different points one and the same flow may be either subsonic or supersonic. However, we can establish a constant reference quantity for a given flow, comparison with which is sufficient to establish the nature of flow. Let us introduce the notation

$$v^* \equiv \left(\frac{\gamma - 1}{\gamma + 1}\right)^{1/2} v_0 \tag{20.17}$$

Now, with the help of (20.9) and (20.15) we rewrite (20.11) as follows:

$$\frac{v^2}{2} + \frac{c^2}{\gamma - 1} = \frac{\gamma + 1}{\gamma - 1}\frac{v_*^2}{2} \tag{20.18}$$

A simple rearrangement of terms in this equation yields

$$\frac{v^2-v_*^2}{2}=\frac{1}{\gamma-1}(v_*^2-c^2) \tag{20.19}$$

It can be seen from this that if $v > v_*$ then $v_* > c$, that is, the flow at the given point is supersonic ($v > c$), and vice versa.

The concepts of the limiting and critical velocities are applicable to the case of any steady isentropic flow, when the strong form of Bernoulli's theorem can be invoked.

Gas Flow in a Heat-Insulated Pipe. We shall now consider the flow of a compressible gas in a long pipe of constant cross section with heat-insulated walls. Since the pipe is long, losses due to viscous friction cannot be neglected. Owing to the heat-insulated walls, the evolved heat is not transmitted to the surrounding medium. Obviously, in such conditions the total energy flux transferred by the gas is conserved. Therefore, from (15.31), we can write

$$\rho v\left(\frac{v^2}{2}+H\right)=\text{constant} \tag{20.20}$$

But if the cross section of the pipe is constant, then ρv = constant; whence

$$\frac{v^2}{2}+H=\text{constant} \tag{20.21}$$

This equation is very like Bernoulli's theorem, but its origin is entirely different. Bernoulli's theorem refers only to isentropic flow. In the present case, however, the entropy of the gas increases due to viscous friction. But owing to the speed of the flow the evolved heat is not transmitted through the walls of the pipe, nor does heat exchange between different bodies of the gas occur. In this sence the conditions recall the Joule-Thomson effect (see (8.56)). But now the velocity of the gas is not completely damped by friction. That is why the expression of the conservation law (20.20) also includes $v^2/2$. Conservation of the quantity $H + v^2/2$ in this case is directly associated with the constancy of the cross section, since in the most general case only the quantity (20.20) is conserved.

The velocity can be eliminated from Eq. (20.21) by replacing it according to the formula $v = q/\rho$, where q is the rate of flow (which is constant in a pipe of constant cross section):

$$\frac{q^2}{2\rho^2}+H=\text{constant} \tag{20.22}$$

Only the thermodynamic quantities H and ρ are involved here.

Let us differentiate Eq. (20.22) with respect to pressure, substituting dH/dp with the help of (8.30):

$$\frac{dH}{dp} = \theta \frac{dS}{dp} + \frac{1}{\rho}$$

For the derivative of entropy with respect to pressure taken along the pipe we obtain

$$\frac{dS}{dp} = -\frac{1}{\rho\theta}\left(1 - v^2 \Big/ \frac{dp}{d\rho}\right) \tag{20.23}$$

where q is replaced by ρv.

Close to the maximum entropy the derivative $dp/d\rho$ is $(\partial p/\partial \rho)_S$. In other words it is equal to c^2. Consequently, entropy reaches the maximum where $v = c$.

The increment of entropy is, according to the second law of thermodynamics, always positive. Therefore, at subsonic flow, when the quantity in parantheses in the right-hand side of (20.23) is positive, the pressure must decrease along the pipe ($dp < 0$). At supersonic flow the pressure increases.

If a supersonic stream is injected into the pipe, it cannot become subsonic inside the pipe, and vice versa, otherwise the entropy would have to decrease spontaneously at some section of the pipe.

All that has been said refers to flows in which all quantities change continuously. However, in very long pipes a discontinuity, or a shock wave, may appear. Suppose a supersonic stream is injected into the same kind of pipe. Since, according to what was just proved, the pressure in the pipe increases, the velocity of the flow through the pipe must gradually decrease. If the velocity decreases to the speed of sound before the gas leaves the pipe the entropy at that point will attain its maximum value. But since it cannot decrease, the steady-state flow must give way to another type of flow, involving a discontinuity.

21

QUASI-ONE-DIMENSIONAL FLOW OF A GAS

Flow in a Pipe of Variable Cross Section. The difference between supersonic and subsonic flow is especially apparent when a gas is moving isentropically through a pipe of slowly varying cross section F. In that case the velocity can be characterized to a good approx-

imation by its mean value over the cross-sectional area of the pipe. If the flow is steady, the discharge of gas through any cross section is the same:

$$\rho v F = \text{constant} \tag{21.1}$$

We take the logarithmic derivative of this expression and get

$$\frac{d\rho}{\rho}+\frac{dv}{v}+\frac{dF}{F}=0 \tag{21.2}$$

We transform the first term as follows:

$$\frac{d\rho}{\rho}=\frac{d\rho}{dp}\frac{dp}{\rho}=\frac{dH}{c^2} \tag{21.3}$$

Here we made use of the condition that the gas is flowing isentropically and $dp/d\rho = (\partial p/\partial\rho)_S$, while $dH = V\,dp = dp/\rho$, since V is the specific volume of the gas. From Bernoulli's equation,

$$dH = -v\,dv \tag{21.4}$$

Substituting the latter two expressions into (21.2), we find

$$\frac{dv}{v}\left(1-\frac{v^2}{c^2}\right)=-\frac{dF}{F} \tag{21.5}$$

Now let the flow be subsonic ($v < c$). Then the expression in the parentheses in the left-hand side of Eq. (21.5) is positive. If the gas is flowing along a convergent pipe, then $dF < 0$. It follows then that $dv > 0$, that is, the flow accelerates. In a diverging pipe subsonic flow decelerates. But if $v > c$, then $(1 - v^2/c^2) < 0$ and the relationship is reversed. Supersonic flow accelerates in a diverging pipe and decelerates in a converging one.

Laval Nozzle. It follows from what has been said that a gas issuing through a convergent orifice from a chamber in which it was at rest cannot attain the speed of sound. For a gas to attain supersonic speeds it must pass through a channel (nozzle) whose cross section first decreases, then reach sonic speed at the minimum cross section, and at last escape through the divergent part of the nozzle, accelerating further. From (21.5), $v = c$ at $dF = 0$, that is, at the narrowest point (the *throat*).

Let us show how to calculate the flow of a gas through a nozzle of given geometry from the equation of state of the gas. For this it is convenient to express the density and enthalpy of the gas in terms of its pressure with the help of the isentropy equation. From Eq. (9.28), for entropy we find that

$$\frac{\rho}{p^{1/\gamma}}=\frac{\rho_0}{p_0^{1/\gamma}} \tag{21.6}$$

Let p_0 and ρ_0 refer to the state of the gas in which it was at rest. We express the enthalpy from (20.3), (20.6), and (20.9), after which we substitute the density ρ, using the isentropy equation (21.6):

$$H=\frac{\gamma}{\gamma-1}\frac{p}{\rho}=\frac{\gamma}{\gamma-1}\left(\frac{p_0^{1/\gamma}}{\rho_0}\right)p^{(\gamma-1)/\gamma} \tag{21.7}$$

From this we determine the density of the flow ρv as a function of the pressure at a given point of the nozzle:

$$\rho v=\rho\,[2\,(H_0-H)]^{1/2}$$
$$=\rho_0\left(\frac{p}{p_0}\right)^{1/\gamma}\left\{\frac{2\gamma}{\gamma-1}\frac{p_0}{\rho_0}\left[1-\left(\frac{p}{p_0}\right)^{1-1/\gamma}\right]\right\}^{1/2}$$

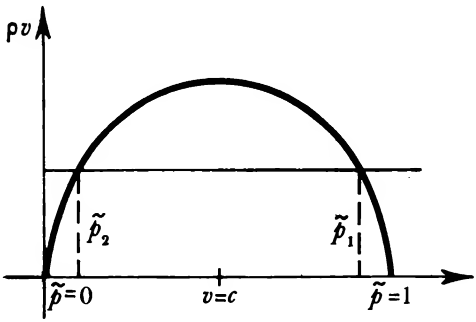


Figure 20

Denoting the ratio p/p_0 as $\tilde{p}$, we reduce the expression for the flow density to the form

$$\rho v=\left(\frac{2\gamma}{\gamma-1}p_0\rho_0\right)^{1/2}\tilde{p}^{1/\gamma}\,(1-\tilde{p}^{1-1/\gamma})^{1/2} \tag{21.8}$$

This expression vanishes both at $\tilde{p}=1$, that is, in the initial state of rest, and at $\tilde{p}=0$, when expanding into vacuum. The maximum is attained at

$$\frac{d}{dp}\rho v=\frac{d\rho}{dp}v-\rho\frac{dH}{dp}\frac{1}{v}=\frac{v}{c^2}-\frac{1}{v}=0 \tag{21.9}$$

that is, at $v=c$. In other words, at this point v equals the local velocity of sound, which corresponds to the nozzle throat.

Let us construct the following curves. First, lay off the quantity $\tilde{p}$ on the abscissa from unity to zero (Figure 20), and ρv on the ordinate. With the help of this graph it is convenient to find the reduced pressure $\tilde{p}$ from the given flow density ρv. Further, represent the cross section F of the nozzle as a function of x from entry ($\tilde{p}=1$) to

exhaust. This is shown graphically in the upper part of Figure 21 as a longitudinal cross section of the nozzle.

Stating a certain x, we determine from the graph the cross section F. Assuming a discharge Q across the whole cross section, we find the flow density, or rate of flow $\rho v = Q/F$. Then, from Figure 20 determine the two values $\tilde{p}_1$ and $\tilde{p}_2$ corresponding to one and the same

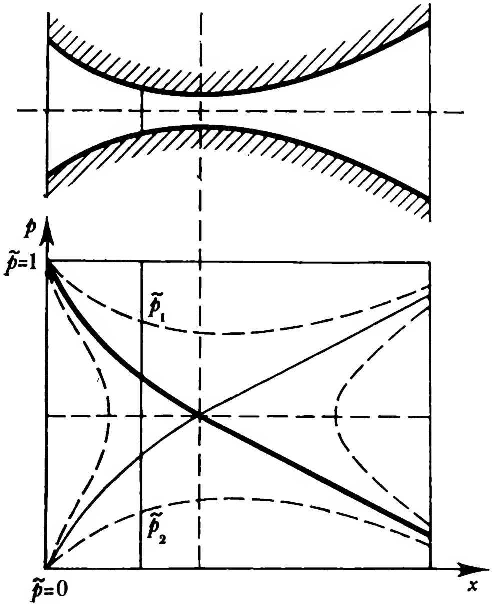


Figure 21

ordinate ρv. Lay them off on the lower part of Figure 21. After joining all the $\tilde{p}_1$ and $\tilde{p}_2$ points by a smooth curve we find that the whole upper curve corresponds to subsonic flow, and the whole lower curve to supersonic flow.

Consequently, the flow Q according to which the curves were drawn does not correspond to the given geometry of the nozzle.

There can exist one, quite definite, value Q_0 at which the point $\tilde{p}_1$ moves along the upper, subsonic, curve before the throat, and in the throat itself it passes over to the lower, supersonic branch. Such a curve is presented in Figure 21 by the heavy line. Note that at the nozzle exit it corresponds to a definite value of pressure $\tilde{p}_0$.

If the pressure at the exit is less than $\tilde{p}_0$, the gas emerges into the surroundings at a higher pressure. Since the efflux is supersonic, this has no effect on the regime inside the nozzle, since the "signals" about

the lower outside pressure cannot enter the nozzle against the stream. The additional expansion of the gas takes place after it leaves the nozzle.

If the external pressure is greater than $\tilde{p}_0$, the expansion of the gas does not follow either of the smooth curves in Figure 21. The efflux regime cannot be continuous. Experience shows that in this case discontinuity transition surfaces, or shock waves, form. Such discontinuities will be examined in general form in Section 25.

The discontinuity surfaces are conical in shape, so that the flow in the nozzle is not one-dimensional, and the quasi-one-dimensional model employed here is not appropriate close to the nozzle exit.

EXERCISES

1. A compressible gas flows into a three-dimensional region from a source located within that region. Determine the minimum size of the source, if its discharge is Q.

Solution. The total gas flow satisfies the equation

$$Q = 4\pi r^2 \rho v = \text{constant}$$

The product ρv has a maximum according to (21.9), which determines the minimum radius for the given Q. The flow outside the source is either wholly subsonic or supersonic.

2. Find the minimum radius of a vortex line in a compressible gas.

Solution. Outside the line the flow is irrotational. If the line is directed along the z axis, the velocity has only the azimuthal component v_φ, and the condition for the flow to be irrotational is

$$\operatorname{curl}_z \mathbf{v} = \frac{1}{r}\frac{d}{dr} r v_\varphi = 0$$

whence

$$v_\varphi = \frac{\Gamma}{2\pi r}$$

where Γ is, as usual, the circulation around the line.

Since the flow is irrotational the strong form of Bernoulli's theory is applicable, namely

$$\frac{1}{2} v_\varphi^2 + H = H_\infty$$

where H_∞ is the enthalpy at an infinite distance from the line at $v_\varphi = 0$. Since $H > 0$, the maximum velocity is $(2H_\infty)^{1/2}$, and the minimum radius is $\Gamma/[2\pi\,(2H_\infty)^{1/2}]$.

22

CHARACTERISTICS OF ONE-DIMENSIONAL NONSTEADY ISENTROPIC FLOW

General Equations. The equations of nonsteady isentropic flow are the most fully studied of the equations of gas dynamics. The isentropic nature of flow should in this case be understood very rigidly: the specific entropy is constant not only for every given particle of the gas but over the whole volume as well. Given these conditions the pressure is a unique function of density, and grad p is replaced by $(\partial p/\partial\rho)_S$ grad $\rho = c^2$ grad ρ. We shall consider the motion to be one-dimensional, as in a pipe. Then the gradient is replaced by $\partial/\partial x$ and the velocity has a projection only on x. The Euler equation (15.11) and the continuity equation (15.6) are reduced to

$$\frac{\partial v}{\partial t}+v\frac{\partial v}{\partial x}=-\frac{c^2}{\rho}\frac{\partial \rho}{\partial x} \tag{22.1}$$

$$\frac{\partial \rho}{\partial t}+v\frac{\partial \rho}{\partial x}+\rho\frac{\partial v}{\partial x}=0 \tag{22.2}$$

The Riemann Invariants. The equations above can be transformed to a form in which the partial derivatives of the quantities involved in the derivatives with respect to x and to t are proportional to one another. For this, multiply the second equation by an undetermined coefficient n and add it to the first to obtain

$$\frac{\partial v}{\partial t}+n\frac{\partial \rho}{\partial t}+(v+n\rho)\frac{\partial v}{\partial x}+\left(\frac{c^2}{\rho}+nv\right)\frac{\partial \rho}{\partial x}=0 \tag{22.3}$$

Furthermore, let us choose the quantity n such that the partial derivatives with respect to t and to x are multiplied by proportional quantities. In other words, we require that

$$\frac{v+n\rho}{1}=\left(\frac{c^2}{\rho}+nv\right)\Big/n \tag{22.4}$$

From this we obtain

$$n^2=\frac{c^2}{\rho^2}$$

or

$$n=\pm\frac{c}{\rho} \tag{22.5}$$

Substitute this expression for n into (22.3) to obtain

$$\frac{\partial v}{\partial t} \pm \frac{c}{\rho}\frac{\partial \rho}{\partial t} + (v \pm c)\frac{\partial v}{\partial x} + \left(\frac{c^2}{\rho} \pm \frac{vc}{\rho}\right)\frac{\partial \rho}{\partial x} = 0$$

After rearranging of terms we find that this equation reduces to

$$\frac{\partial v}{\partial t} \pm \frac{c}{\rho}\frac{\partial \rho}{\partial t} + (v \pm c)\left(\frac{\partial v}{\partial x} \pm \frac{c}{\rho}\frac{\partial \rho}{\partial x}\right) = 0 \tag{22.6}$$

Thus, the partial derivatives are really proportional for any choice of sign:

$$\partial v + \frac{c\,\partial\rho}{\rho} \quad \text{and} \quad (v+c)\left(\partial v + \frac{c\,\partial\rho}{\rho}\right)$$

or

$$\partial v - \frac{c\,\partial\rho}{\rho} \quad \text{and} \quad (v-c)\left(\partial v - \frac{c\,\partial\rho}{\rho}\right)$$

But these equations can be written more simply if we make use of the fact that if a process is totally isentropic the thermodynamic quantities are expressed in terms of one of them, in the present case the density, ρ. Introducing the quantity

$$u \equiv \int c\,\frac{d\rho}{\rho} \tag{22.7}$$

we can replace derivatives with respect to ρ by derivatives with respect to u:

$$\frac{c}{\rho}\frac{\partial \rho}{\partial t} = \frac{\partial u}{\partial t}, \qquad \frac{c}{\rho}\frac{\partial \rho}{\partial x} = \frac{\partial u}{\partial x} \tag{22.8}$$

after which Eqs. (22.6) acquire a very symmetrical form:

$$\frac{\partial}{\partial t}(v \pm u) + (v \pm c)\frac{\partial}{\partial x}(v \pm u) = 0 \tag{22.9}$$

In both terms the quantities under the derivative sign are the same. Take for example the first of these equations and rewrite it in the form

$$-\frac{\partial (v+u)/\partial t}{\partial (v+u)/\partial x} = v + c \tag{22.10}$$

It will be observed that on the left we have the derivative $\partial x/\partial t$ at constant value of $v + u$. Thus if in the x,t-plane we define a curve by

$$\frac{dx}{dt} = v + c \tag{22.11}$$

the quantity $v + u$ remains constant along it. Similarly, along the other curve satisfying the equation

$$\frac{dx}{dt} = v - c \tag{22.12}$$

the value $v - u$ is constant. The invariants $v \pm u$, discovered by Georg F. B. Riemann, the founder of gas dynamics, bear his name.

Characteristics. The curves described by Eqs. (22.11) and (22.12) are known as the characteristics of equations of gas dynamics. Let us now explain their meaning. Let at some initial instant a small disturbance affect an arbitrary particle of a gas. This disturbance will spread in both directions with a speed $\pm c$ with respect to the gas, in the most general case c being a variable quantity. But since

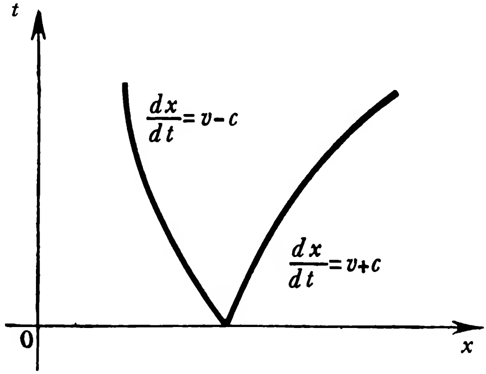


Figure 22

the gas particles themselves travel with a velocity v with respect to a stationary reference frame, the disturbances will propagate in that frame with a velocity $v \pm c$.

The characteristics passing through a certain point show how the disturbances, or "signals", emanating from that point propagate (Figure 22). The state of the gas at that point affects the state of the gas between the two characteristics, but it does not affect the gas particles outside this domain. There is a profound analogy here with the light cone in electrodynamics [Sec. 13]. The fact that the velocity of sound is variable and is added to the velocity of the gas substantially complicates the picture and, as will be shown later in Section 23, leads to the appearance of discontinuities, or shock waves, of which there is no analogue in electrodynamics.

The invariant $v + u$ is conserved along the characteristic (22.11), and the invariant $v - u$ along the characteristic (22.12). From this we can see in general form how to effect the solution of the system (22.1)-(22.2) or (22.9). Let the state of a gas be given along a curve AB in the x, t-plane (Figure 23). Segment AB is everywhere directed so that both characteristics through any of its points make a larger angle with the x axis than the segment at that point.

By analogy with the term accepted in relativity theory, such a segment AB is called spacelike [Sec. 13].

Since v and u (and with them c, ρ, p) are known on the segment AB, the initial segments of the characteristics of both families $v \pm c$ can be drawn through every point. Take points *1* and *2* and the point *3* where the line $dx/dt = v + c$ drawn from point *1* intersects with line $dx/dt = v - c$ drawn from point *2*. Of course, both segments, *13*

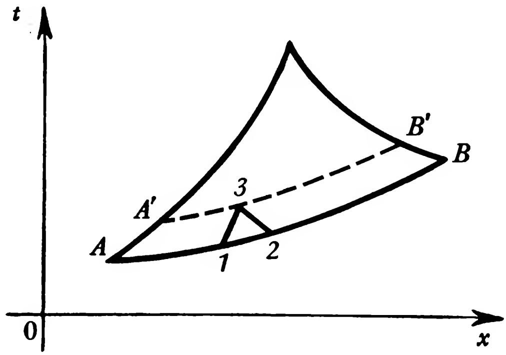


Figure 23

and *23*, are assumed sufficiently small and are drawn as straight lines. But along line *13* the invariant $v + u$ is conserved, and so is the invariant $v - u$ along *23*. Therefore we obtain the following two equations:

$$v_3 + u_3 = v_1 + u_1 \tag{22.13}$$

$$v_3 - u_3 = v_2 - u_2 \tag{22.14}$$

which completely define the state at point *3*.

The points *3* form a smooth curve, shown by the dashed line. Along this curve the state of the gas is known again from Eqs. (22.13), and this holds up to the apex C of the curvilinear triangle. In this triangle the state of the gas can be determined from its state on line AB.

We can see from the construction why segment AB must lie below both characteristics *12* and *13*, as otherwise the required intersection point *3* would be lacking.

It is possible to state the initial functions on two intersecting timelike segments as well (Figure 24). The construction of the characteristics can be seen from the diagram: *3* is determined from *1* and *2*; *6* from *3* and *4*; *7* from *3* and *5*, etc.

The initial state defined on segment AB (Figure 23) and on segments AC and BC (Figure 24), so to say, propagates through the gas.

The propagation velocity with respect to the gas is always equal to $\pm c$. Accordingly, the equations of gas dynamics belong to the

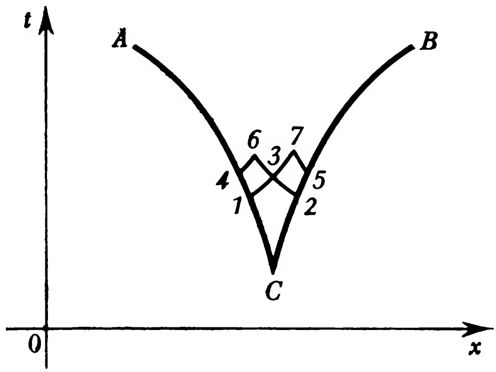


Figure 24

category of wave (but nonlinear) equations. The description of propagating processes relates them with the wave equations of electrodynamics.

Propagation of Weak Discontinuities. The functions v and u defined on segment AB need not necessarily have the same analytical form along the whole segment. At some points their derivatives may suffer a discontinuity, provided there is no discontinuity in the functions themselves. Indeed, the last condition is sufficient for the characteristic equations (22.11) and (22.12) through the point of discontinuity of the derivative to have one value. The Riemann invariants $v \pm u$ on these characteristics must also be stated uniquely, but their derivatives have a discontinuity when going over to neighbouring characteristics. The solution not analytical at point *1* on segment AB has discontinuities in both characteristics passing through this point. But these discontinuities refer only to the derivatives, and not to the quantities themselves. They are therefore called *weak discontinuities*, as distinct from strong discontinuities, or shock waves (see Sec. 25). The condition of continuity of the functions themselves is necessary for the first-order equations (22.1) and (22.2) to have meaning.

23

SIMPLE WAVES

Special Solutions. In the most general case a solution of the system of two partial differential equations (22.1) and (22.2) of the first order contains two arbitrary functions. But there exists an important class of special solutions containing one arbitrary function, which can be obtained from the conditions of the problem.

To construct such solutions let us write the set (22.9) for each of the invariants $v+u$ and $v-u$ separately:

$$\frac{\partial}{\partial t}(v+u)+(v+c)\frac{\partial}{\partial x}(v+u)=0 \tag{23.14}$$

$$\frac{\partial}{\partial t}(v-u)+(v-c)\frac{\partial}{\partial x}(v-u)=0 \tag{23.15}$$

In the most general case these equations are interrelated, as c depends upon u. But if, for example, we put $v-u$ equal to a constant quantity in a certain domain of the flow, Eq. (23.15) will be automatically satisfied. Having stated this constant, we can express u and c as functions of v and substitute them into Eq. (23.14), which will thus contain only one unknown function, v. The solution of the set is expressed in terms of one arbitrary function, and it is called a *simple wave.*

Let us construct this solution. The equation of the characteristic defined by (23.14) is Eq. (23.11):

$$\left(\frac{\partial x}{\partial t}\right)_{v+u=\text{constant}}=v+c \tag{23.16}$$

In the most general case the derivative dx/dt is taken for a constant Riemann invariant $v+u$. But in a simple wave u is a function of v. Consequently, c is also a function of v; hence $v+u$ and $v+c$ are functions only of v. Therefore (23.16) can be written as follows:

$$\left(\frac{\partial x}{\partial t}\right)_{v=\text{constant}}=v+c(v) \tag{23.17}$$

This equation can be solved directly:

$$x=[v+c(v)]\,t+f_{+}(v) \tag{23.18a}$$

where f_+ is an arbitrary function of the velocity. Examples of determining it from problem conditions will be given later.

In the domain of a simple wave, where $v-u=$ constant, all characteristics of the family (23.18*a*) are rectilinear. To each line corresponds a definite value of the velocity v, defining the slope $v+$ $+c(v)$ and the segment $f_+(v)$ which is cut off on the x axis at $t=0$.

The other simple wave we obtain from the solution $v + u =$ constant. It has a rectilinear family of characteristics:

$$x = [v - c(v)]\, t + f_-(v) \tag{23.18b}$$

Every simple wave also contains a family of curvilinear characteristics. At $v - u =$ constant it is described by the equation $dx/dt = v - c$, and at $v + u =$ constant by the equation $dx/dt = v + c$.

Criterion for the Appearance of a Simple Wave. There exists a simple criterion for the involvement of a simple wave in the solution of a problem of gas dynamics. Let some region of a flow be unaf-

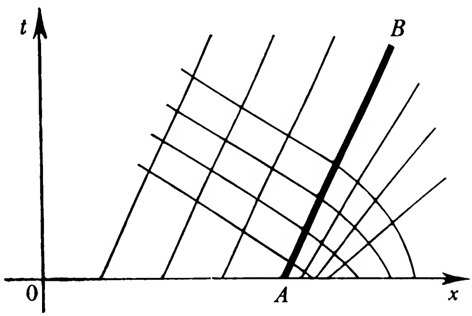


Figure 25

fected by disturbances. In it $v =$ constant and $u =$ constant. In such a flow region both families of characteristics have the form of parallel straight lines (Figure 25). Characteristic AB is on the extreme right in this family. Further to the right lies the disturbed region of the family. Hence, line AB describes the propagation of a weak discontinuity separating the steady flow from the disturbed, where the function possesses a different analytical form.

The continuations of the other family of characteristics intersecting AB are curvilinear in the region of nonsteady flow (Figure 25). In the steady-flow region each of them corresponds to the same value of $v + u$, going with it to the domain to the right of AB. But in the most general case $v + u$ is conserved along this family of characteristics. Hence $v + u$ is constant over the whole domain of disturbed flow and not only over the characteristics of the corresponding family. The quantity $v + u$ is carried by the characteristics from the steady-flow domain. In other words, to the right of AB there must evidently be a simple wave.

But then the family of characteristics which to the left of AB are rectilinear and parallel to AB consists of straight lines also in the domain to the right of AB. Only there the characteristics of

this family are differently inclined. Thus, on the boundary of steady flow or vacuum there is always a simple wave.

Simple Waves in Gases With a Constant Adiabatic Exponent. The formulas for simple waves have an especially convenient form for gases with a constant adiabatic exponent. Since pressure is proportional to ρ^γ (see (21.6)), $c = [(\partial p/\partial \rho)_S]^{1/2}$ is proportional to $\rho^{(\gamma-1)/2}$. Consequently

$$\frac{\gamma-1}{2}\frac{d\rho}{\rho}=\frac{dc}{c}$$

or

$$u=\int c\,\frac{d\rho}{\rho}=\frac{2}{\gamma-1}\int c\,\frac{dc}{c}=\frac{2c}{\gamma-1} \tag{23.19}$$

For example, for air, where $\gamma = 7/5$, we obtain $u = 5c$. In a simple wave of the type (23.18*a*) we can put

$$v - u = \text{constant} = -u_0$$

or

$$c=\frac{v+u_0}{5}$$

whence

$$x=\left(\frac{6}{5}v+\frac{u_0}{5}\right)t+f_+(v) \tag{23.20}$$

where u_0 is the value of u in the region where the air is at rest.

A curious case is that of $\gamma = 3$. An ideal gas cannot, of course, have such an adiabatic exponent, but formulas involving $\gamma = 3$ are necessary for constructing a general solution of the equations of gas dynamics in Section 24. Besides, the pressure of the dense detonation products of explosives of the type of TNT approximately follow the law $p = A\rho^3$.

At $\gamma = 3$ we obtain $u = c$. But then the system of equations (23.14) and (23.15) separates completely. Its most general solution has the form

$$x = (v + c)\,t + f + (v + c) \tag{23.21a}$$

$$x = (v - c)\,t + f - (v - c) \tag{23.21b}$$

The waves travelling in either direction propagate independently and are not mutually disturbed. At $\gamma \neq 3$ only simple waves do not produce any waves in the opposite direction.

If the functions f_+ and f_- are known from the problem conditions, Eqs. (23.21*a*, *b*) fully define in implicit form the flow velocity and the velocity of sound, and hence the other thermodynamic quantities, as functions of position and time.

Rarefaction and Compression Waves. Suppose a velocity distribution of the type shown in Figure 26 is produced at some initial instant $t = 0$ in a stationary gas of homogeneous thermodynamic state. The velocity increases everywhere along the x axis from point *1* until it reaches some maximum value at point *2*, after which it decreases and becomes zero at point *3*. The resulting flow adjoins on a steady flow, that is, a region of rest, and therefore represents a simple wave. From Eq. (23.12)

$$x = (v + c)\, t + f_+ (v)$$

where the function f_+ is determined by the initial velocity distribution over the section *1-3*. This simple wave consists of two parts.

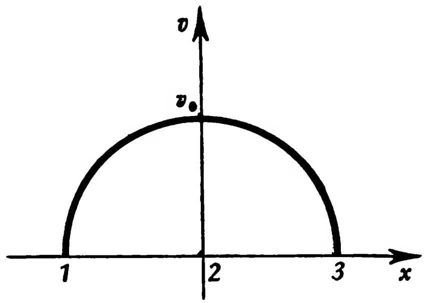


Figure 26

The gas between *1* and *2* expands because the particles move the faster the farther they are from point *1*. This section of the flow is called a *rarefaction wave.*

The particles at point *2* and to the right of it catch up with those still farther to the right. As a result the gas compresses. This is a *compression wave*. The properties of rarefaction waves and compression waves are very different.

To see this let us find the steepness of the waves at some instant, that is, the derivative $(\partial v/\partial x)_t$ by the rule of differentiating implicit functions:

$$\left(\frac{\partial v}{\partial x}\right)_t = \left[\left(1 + \frac{dc}{dv}\right) t + \frac{df_+}{dv}\right]^{-1} \tag{23.22}$$

In the region of the rarefaction wave the derivative df_+/dv is positive constant; dc/dv for the given simple wave ($v - u =$ constant) is also positive and constant. Therefore expression in brackets in (23.22) cannot vanish anywhere. As t increases the steepness of the wave decreases and the rarefaction wave extends in space.

In the compression wave the derivative df_+/dv is, in accordance with the velocity profile in Figure 26, negative. Therefore there in-

evitably comes a moment t' when the steepness of the front becomes infinite (curve *2* in Figure 27). If we formally continue the solution up to that instant, the distribution curve will "curl over", like a breaker rolling up a beach. But obviously the velocity cannot have several values at the same point. Therefore in real conditions after the vertical tangent to the velocity profile appears, a shock wave develops (the straight segment *3* in Figure 27).

The reason why the vertical tangent appears can be understood as follows. The characteristics coming from the points where the

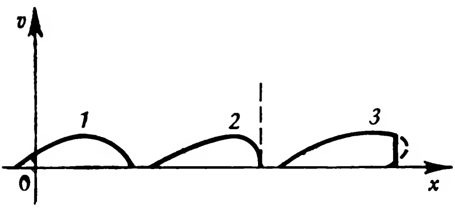


Figure 27

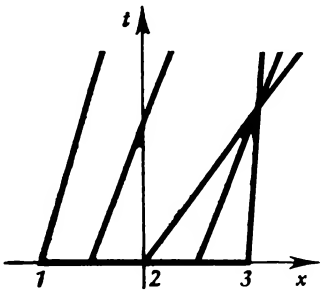


Figure 28

velocity is greater have, from the equation $(dx/dt) = v + c$, a greater inclination to the x axis. They possess greater v. Moreover, they come from a domain where the compression is greater, consequently their c is also greater.

Because of this the characteristics to the left overtake the characteristics to the right. But it is physically impossible for characteristics to intersect, since they carry different values of the Riemann invariants. If two characteristics of the same family were to intersect, the corresponding quantity $v \pm u$ would have two different values at the same point in space and at the same moment of time, which is impossible.

In a rarefaction wave the characteristics fan out (Figure 28) and therefore cannot intersect.

EXERCISES

1. A long cylinder is divided in two by a partition: one half (at $x \leqslant 0$) contains a homogeneous stationary gas with a constant adiabatic exponent γ, the other (at $x > 0$) contains a vacuum. The partition is removed instantaneously at time $t = 0$. Describe the motion of the gas.

Solution. A simple wave (23.18*b*) appears on the boundary of the stationary gas and travels to the left. Since the problem conditions do not involve quantities possessing the dimension of length, the function f_- involved in the expression for x can in the most general case be equal only to zero. Consequently

$$x=(v-c)\,t=\left(v-\frac{\gamma-1}{2}u\right)t$$

In this simple wave the sum $v+u$ is constant: $v+u=u_0=2c_0/(\gamma-1)$. At the boundary of the gas with the region of rest, $v=0$ and $x=-c_0t$. At the boundary with the vacuum, $u=0$, $c=0$, and $v=2c_0/(\gamma-1)$. Expansion into vacuum takes place with a velocity $2c_0/(\gamma-1)$, so that the boundary with the vacuum is given by the equation $x=2c_0t/(\gamma-1)$. (This can be compared with the speed of steady scattering in vacuum $c_0\,[2/(\gamma-1)]^{1/2}$. At an arbitrary point

$$x=\left(\frac{1+\gamma}{2}v-c_0\right)t$$

All the quantities depend solely on the ratio x/t, hence with time the simple wave domain expands while remaining similar to itself. Such a wave is said to be *automodel* or *self-similar*. All characteristics of the family $(dx/dt)=v-c$ fan out of the origin of the coordinate system.

2. Find the equations of curvilinear characteristics in a self-similar simple wave.

Solution. The differential equation of these curves is written as follows:

$$\frac{dx}{dt}=v+c$$

From Exercise 1 we find:

$$v=\frac{2}{1+\gamma}\left(\frac{x}{t}+c_0\right)$$

$$c=\frac{2}{1+\gamma}c_0-\frac{\gamma-1}{\gamma+1}\frac{x}{t}$$

Consequently

$$\frac{dx}{dt}=\frac{4c_0}{1+\gamma}+\frac{3-\gamma}{1+\gamma}\frac{x}{t}$$

This homogeneous equation is integrated by substituting $x=wt$:

$$t\frac{dw}{dt}=\frac{4c_0}{1+\gamma}-\frac{2(\gamma-1)w}{1+\gamma}$$

From this it is easy to see the following particular solution:

$$w_0=\frac{2c_0}{\gamma-1}$$

The general solution has the form

$$\frac{x}{t}=w=\frac{2c_0}{\gamma-1}+At^{-2(\gamma-1)/(\gamma+1)}$$

where A is an integration constant defining a characteristic in the given family.

3. A piston moves out of a cylinder at a constant velocity v_0. Describe the motion of the gas behind the piston.

Solution. The solution is obtained from Exercise 1. If the velocity of the piston, v_0, is less than $2c_0/(\gamma - 1)$, a simple wave region forms between it and the stationary gas. At the piston itself $v = v_0$, so that the velocity of sound is equal to $c_0 - v_0 (\gamma - 1)/2$. From this the density and pressure of the gas at the piston can be found with the help of the isentropy equation.

Between the piston and the simple wave lies a region of steady flow. Indeed, if the velocity of the piston is v_0 the rectilinear characteristic farthest to the right is described by the equation $(dx'/dt) = v_0 - c'$, where c' is the velocity of sound at the piston. It can be determined from the equation $v_0 + 2c'/(\gamma - 1) = 2c_0/(\gamma - 1)$. At $v = v_0$ we obtain $v_0 - c' = (\gamma + 1) v_0/2 - c_0$. Substituting $v_0 - c'$ in the equation of the extreme right characteristic, we obtain $x' = [(\gamma + 1) v_0/2 - c_0]\, t$. It lags behind the piston because $x_p - x' = v_0 t - [(\gamma + 1) v_0/2 - c_0]\, t$ and c_0 is by definition greater than $(\gamma - 1) v_0/2$.

At $v_0 > 2c_0/(\gamma - 1)$ a self-similar simple wave is formed completely, that is, it contains the whole fan of rectilinear characteristics. A region of vacuum develops between the piston and the gas. Of course, this is true only on the assumption that the gas forms a continuous medium, without account of the velocity distribution over the molecules.

4. A piston moves out of a cylinder filled with air ($\gamma = 7/5$) according to the law $x = x_0(t)$, but so that at $t = 0$, $x_0 = 0$ and $\dot{x}_0 = 0$. Determine the motion of the air.

Solution. From the general solution (23.18*b*),

$$x = (v - c)\, t + f_-(v)$$
$$v + 5c = 5c_0$$

whence

$$x = \left(\frac{6}{5} v - c_0\right) t + f_-(v)$$

On the piston $v = \dot{x}_0(t)$, therefore

$$x_0(t) - \left(\frac{6}{5}\dot{x}_0(t) - c_0\right) t = f_-(\dot{x}_0(t))$$

Thereby the function f_- is defined in parametric form: the argument $\dot{x}_0(t)$, or v, and the value of the function itself are known for every value of t. The condition $\dot{x}_0(0) = 0$ was imposed to prevent the appearance of a self-similar simple-wave region at the initial moment of motion, as in Exercise 3.

5. Show that if a piston is moving in a cylinder with uniform acceleration according to the equation $x = -at^2/2$, then after a certain time a point will appear where $(\partial v/\partial x)_t$ becomes infinitely great.

Solution. From Exercise 4, the function f_- is determined by the equation

$$f_-(-at) = -\frac{1}{2}at^2 + \frac{6}{5}at^2 + c_0 t = \frac{7}{10}at^2 + c_0 t = \frac{7}{10}\frac{(at)^2}{a} + \frac{c_0(at)}{a}$$

whence

$$f_-(v) = \frac{7}{10}\frac{v^2}{a} - \frac{c_0 v}{a}$$

Substitute this expression into the simple wave equation to get

$$x = \left(\frac{6}{5}v - c_0\right)t + \frac{7}{10}\frac{v^2}{a} - \frac{c_0 v}{a}$$

Solving the quadratic equation with respect to v, we obtain

$$v = \frac{6at - 5c_0 - [(6at - 5c_0)^2 + 70a^2(x + c_0 t)]^{1/2}}{7}$$

The minus before the brackets results from the fact that at $x = -c_0 t$ the velocity must become zero: up to that point a simple compression wave propagates through stationary air. The derivative $(\partial v/\partial x)_t$ is equal to

$$\left(\frac{\partial v}{\partial x}\right)_t = -\frac{5}{[(6at - 5c_0)^2 + 70a^2(x + c_0 t)]^{1/2}}$$

In a simple wave the second term in the brackets is always nonnegative: $x \geqslant -c_0 t$. Consequently, the denominator becomes zero only when both terms are zero, that is, at $x = -c_0 t$ and $t = 5c_0/(6a)$. In this case the vertical tangent appears at the foremost point of the wave. It sometimes also appears in the centre of a simple compression wave.

If the piston were moving with uniform acceleration out of a pipe, the expression in the brackets would be $(6at + 5c_0)^2 + 70a^2(x + c_0 t)$, which does not vanish at $t > 0$, $|x| < c_0 t$.

24

ONE-DIMENSIONAL NONSTEADY ISENTROPIC FLOW: INTERACTION OF SIMPLE WAVES

In this section we shall seek the general solutions of the equations of one-dimensional nonsteady flow involving two arbitrary functions of position and time. But first we must make a brief digression of a mathematical nature.

Transformation of Independent Variables. The system of equations (23.14) and (23.15) is reduced to a linear one if x and t are made de-

pendent variables, and v and u, independent. In a number of very important cases this enables an easy solution of the system.

To go over to the new variables the following device is convenient. Let it be necessary to go over from the old variables x, t to new variables v, u. Then the volume element $dx\,dt$ transforms as follows:

$$dx\,dt = \frac{\partial(x, t)}{\partial(v, u)}\,dv\,du \tag{24.1}$$

where the expression in front of $dv\,du$ is the *functional determinant*, or *Jacobian*:

$$\frac{\partial(x, t)}{\partial(v, u)} \equiv \left(\frac{\partial x}{\partial v}\right)_u \left(\frac{\partial t}{\partial u}\right)_v - \left(\frac{\partial x}{\partial u}\right)_v \left(\frac{\partial t}{\partial v}\right)_u \tag{24.2}$$

Taking, for example, a transformation from Cartesian coordinates x, y to polar coordinates r, φ, we find that the Jacobian is equal to r, as it should be.

The fraction notation of the Jacobian is explained as follows.

Suppose that the transformation is first effected from x, t to v, u, and then from v, u to z, w. Then the following transformation equations must be written:

$$dx\,dt = \frac{\partial(x, t)}{\partial(v, u)}\,dv\,du, \qquad dv\,du = \frac{\partial(v, u)}{\partial(z, w)}\,dz\,dw$$

$$dx\,dt = \frac{\partial(x, t)}{\partial(v, u)}\,\frac{\partial(v, u)}{\partial(z, w)}\,dz\,dw \tag{24.3}$$

But we can go over from x, t to z, w directly:

$$dx\,dt = \frac{\partial(x, t)}{\partial(z, w)}\,dz\,dw \tag{24.4}$$

Comparing (24.3) with (24.4), we find that

$$\frac{\partial(x, t)}{\partial(v, u)} \times \frac{\partial(v, u)}{\partial(z, w)} = \frac{\partial(x, t)}{\partial(z, w)} \tag{24.5}$$

Thus, the symbol $\partial(v, u)$ so to say cancels out, as in fractions. Accordingly, if necessary, it can be legitimately added to the expression for a Jacobian.

For transformations it is sometimes convenient to write ordinary partial derivatives as Jacobians:

$$\frac{\partial(u, t)}{\partial(x, t)} = \left(\frac{\partial u}{\partial x}\right)_t \left(\frac{\partial t}{\partial t}\right)_x = \left(\frac{\partial u}{\partial x}\right)_t \tag{24.6}$$

because the second term of the determinant involves $(\partial t/\partial x)_t = 0$.

Note also that a permutation of any pair of variables x, t or v, u is equivalent to reversing the sign of the Jacobian, as of any determinant.

Let us apply the relationships obtained to the set of equations (23.14) and (23.15). Multiply each equation by $\partial(x, t)/\partial(v, u)$.

The derivatives involved in the equations transform in the process as follows:

$$\frac{\partial (x,\ t)}{\partial (v,\ u)}\left(\frac{\partial v}{\partial t}\right)_x=\frac{\partial (x,\ t)}{\partial (v,\ u)}\frac{\partial (v,\ x)}{\partial (t,\ x)}=-\frac{\partial (v,\ x)}{\partial (v,\ u)}=-\left(\frac{\partial x}{\partial u}\right),$$

where all the mentioned properties of Jacobians were utilized.

After going over to the independent variables v, u, the system takes the form (the subscripts have been deleted since they are now self-evident):

$$-\frac{\partial x}{\partial u}+\frac{\partial x}{\partial v}+(v+c)\left(\frac{\partial t}{\partial u}-\frac{\partial t}{\partial v}\right)=0 \tag{24.7}$$

$$-\frac{\partial x}{\partial u}-\frac{\partial x}{\partial v}+(v-c)\left(\frac{\partial t}{\partial u}+\frac{\partial t}{\partial v}\right)=0 \tag{24.8}$$

Adding and subtracting these equations, we obtain a simpler set:

$$-\frac{\partial x}{\partial u}+v\frac{\partial t}{\partial u}-c\frac{\partial t}{\partial v}=0 \tag{24.9}$$

$$\frac{\partial x}{\partial v}+c\frac{\partial t}{\partial u}-v\frac{\partial t}{\partial v}=0 \tag{24.10}$$

Let us go over from the thermodynamic variable u to the enthalpy H. Since the derivatives are taken at constant entropy, we can write

$$dH=\frac{dp}{\rho}=\frac{1}{\rho}\frac{dp}{d\rho}d\rho=\frac{c^2}{\rho}d\rho=c\,du \tag{24.11}$$

Then the system (24.9)-(24.10) takes the form

$$-\frac{\partial x}{\partial H}+v\frac{\partial t}{\partial H}-\frac{\partial t}{\partial v}=0 \tag{24.12}$$

$$\frac{\partial x}{\partial v}+c^2\frac{\partial t}{\partial H}-v\frac{\partial t}{\partial v}=0 \tag{24.13}$$

Solution of the Basic System of Equations. The obtained set of two linear equations of the first order are conveniently transformed to one second-order equation. A method similar to the one used for the transformation from the Maxwell's equations [12.34] and [12.35] for fields to the wave equations for potentials is applicable. In the event the first pair of Maxwell's equations is satisfied identically. Let us put

$$x=v\frac{\partial \chi}{\partial H}-\frac{\partial \chi}{\partial v} \tag{24.14}$$

$$t=\frac{\partial \chi}{\partial H} \tag{24.15}$$

Here the function χ is similar to a potential. Then

$$\frac{\partial x}{\partial H}=v\frac{\partial^2 \chi}{\partial H^2}-\frac{\partial^2 \chi}{\partial H\,\partial v},\qquad \frac{\partial t}{\partial H}=\frac{\partial^2 \chi}{\partial H^2},\qquad \frac{\partial t}{\partial v}=\frac{\partial^2 \chi}{\partial v\,\partial H}$$

so that Eq. (24.12) is carried out identically. Equation (24.13) reduces to the form

$$c^2 \frac{\partial^2 \chi}{\partial H^2} - \frac{\partial^2 \chi}{\partial v^2} + \frac{\partial \chi}{\partial H} = 0 \tag{24.16}$$

In this form it is very like a wave equation. To solve it we must state the dependence of c upon H. For gases with a constant adiabatic exponent we have, from (20.9): $c^2 = (\gamma - 1) H$. Therefore (24.16) reduces to the following final form:

$$(\gamma - 1) H \frac{\partial^2 \chi}{\partial H^2} - \frac{\partial^2 \chi}{\partial v^2} + \frac{\partial \chi}{\partial H} = 0 \tag{24.17}$$

Fortunately, the equation has a very simple solution precisely for monatomic and diatomic gases, that is, for $\gamma = 5/3$ and $\gamma = 7/5$. We write

$$\gamma \equiv \frac{3+2n}{1+2n} \tag{24.18}$$

where n is a positive integer or zero. Then $n = 1$ yields $\gamma = 5/3$, and $n = 2$ yields $\gamma = 7/5$.

For a given n denote the solution χ_n. Then

$$\frac{2}{2n+1} H \frac{\partial^2 \chi_n}{\partial H^2} - \frac{\partial^2 \chi_n}{\partial v^2} + \frac{\partial \chi_n}{\partial H} = 0 \tag{24.19}$$

Differentiate (23.19) with respect to H. After a simple rearrangement of terms we obtain a similar equation for $\partial \chi_n / \partial H$:

$$\frac{2}{2n+3} H \frac{\partial^2}{\partial H^2} \frac{\partial \chi_n}{\partial H} - \frac{2n+1}{2n+3} \frac{\partial^2}{\partial v^2} \frac{\partial \chi_n}{\partial H} + \frac{\partial}{\partial H} \frac{\partial \chi_n}{\partial H} = 0 \tag{24.20}$$

If we make the substitution

$$v \left(\frac{2n+3}{2n+1} \right)^{1/2} \equiv v' \tag{24.21}$$

then $\partial \chi_n / \partial H$ satisfies the same kind of equation as $\chi_{n+1}(v', H)$:

$$\chi_{n+1}(v', H) = \frac{\partial \chi_n (v, H)}{\partial H} \tag{24.22}$$

Discarding the prime at v on the left, and expressing v on the right as v', we arrive at the recurrent formula

$$\chi_{n+1}(v, H) = \frac{\partial}{\partial H} \chi_n \left(v \left[\frac{2n+1}{2n+3} \right]^{1/2}, H \right) \tag{24.23}$$

At $n = 0$ we have the mentioned case of $\gamma = 3$ for which Eqs. (23.14) and (23.15) had particularly simple solutions. But here a solution in a different form will be required. Write (24.19) for $n = 0$, and go back from H to c according to the formula

$$H = \frac{c^2}{\gamma - 1} = \frac{c^2}{2}$$

Substitution of c for H yields

$$c^2 \frac{1}{c} \frac{\partial}{\partial c} \frac{1}{c} \frac{\partial \chi_0}{\partial c} - \frac{\partial^2 \chi_0}{\partial v^2} + \frac{1}{c} \frac{\partial \chi_0}{\partial c} = 0$$

which reduces to the standard wave equation

$$\frac{\partial^2 \chi_0}{\partial c^2} - \frac{\partial^2 \chi_0}{\partial v^2} = 0 \tag{24.24}$$

the solution of which has the form

$$\begin{aligned} \chi_0 &= \chi_{01}(c+v) + \chi_{02}(c-v) \\ &= \chi_{01}(\sqrt{2H}+v) + \chi_{02}(\sqrt{2H}-v) \end{aligned} \tag{24.25}$$

(see (18.12)).

This solution involves two arbitrary functions and is therefore a general solution. The quantities $\chi_1, \chi_2, \ldots$ are determined from it by differentiating and also involve two arbitrary functions.

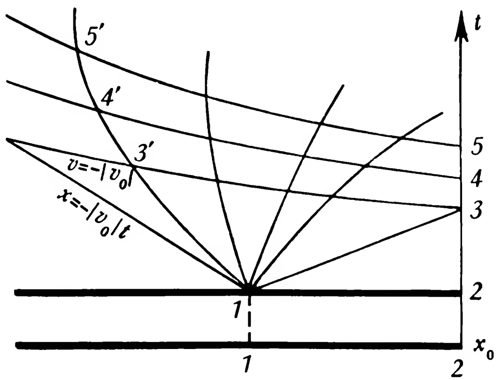


Figure 29

The Meaning of the Obtained Solution. We shall now show the applications of solution (24.25) and similar ones at $n \geqslant 0$. It was shown in the previous section that a simple wave appears at a flow boundary with a region of constant v and u. The solution obtained in this section is inapplicable to a simple wave, because here v and u are independent variables, while in a simple wave $v \pm u$ = constant. Accordingly, solution (24.25) and solutions derived from it hold only in flow regions that do not border on steady flow. Such a solution can border only on a wall or a simple wave.

Let us show how this occurs (Figure 29). Let us consider a pipe sealed at one end and containing a gas separated from a vacuum by a partition *1*. If the partition is removed, a self-similar simple wave with a fan of rectilinear characteristics appears. The equation of

the characteristic on the far right is $-x = c_0 t$. At some instant the characteristic reaches the wall *2* producing a picture of simple wave reflection. A curvilinear characteristic *3-3'* passes to the left of the wall, followed by a series of similar characteristics *4-4'*, etc. To the right of line *3-3'* lies the domain described by the solution obtained in this section.

Imagine that the pipe is continued beyond point *2* by a section symmetrical to it up to a point at a distance *2-1* to the right. If the

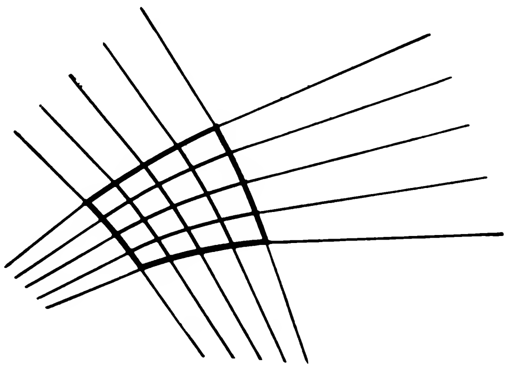


Figure 30

partitions on both sides are removed simultaneously, a wave identical to the simple wave emanating from point *1* will travel to the left towards it. A region of interaction of simple waves appears, lying between line *3-3'* on the left and a symmetrically located line to the right. Thus the reflection from the wall is equivalent to an interaction with a head-on simple wave that is in mirror symmetry with respect to the given wave.

A more complex interaction picture is presented in Figure 30, in which both pencils of rectilinear characteristics are represented. In the intersection domain the characteristics of both waves are curved. Outside the domain they are rectilinear.

Boundary Conditions for the General Solution. To determine the flow in the region of interaction of simple waves the conditions at the boundary of the region must be used. This may be either on the wall or on the boundary with a simple wave.

On a fixed wall the velocity of the gas is zero. Hence, from (24.14) the coordinate of the wall x_0 is equal to

$$x_0 = -\left(\frac{\partial \chi}{\partial v}\right)_{v=0} \tag{24.26}$$

On the boundary with a simple wave both the wave equations

$$x = (v \pm c)\,t + f_{\pm}(v)$$

and equations (24.14) and (24.15) must be satisfied. Substituting these expressions into the equation of a simple wave we obtain

$$v\,\frac{\partial \chi}{\partial H} - \frac{\partial \chi}{\partial v} = (v \pm c)\,\frac{\partial \chi}{\partial H} + f_{\pm}(v)$$

But with the help of equation (24.11) dH/c can be replaced by du:

$$-\frac{\partial \chi}{\partial v} = \pm\,\frac{\partial \chi}{\partial u} + f_{\pm}(v)$$

In a simple wave one of the two relationships, $v \pm u =$ constant, is always satisfied so that

$$\frac{du}{dv} = \pm 1$$

Hence, the condition imposed on χ has the form

$$-\left(\frac{du}{dv}\,\frac{\partial \chi}{\partial u} + \frac{\partial \chi}{\partial v}\right) = -\frac{d\chi}{\partial v} = f_{\pm}(v)$$

Finally we arrive at the following condition:

$$\chi = -\int f_{\pm}(v)\,dv \tag{24.27}$$

At the boundary with a self-similar simple wave (see Exercise 1, Sec. 22) for which $f_{\pm} = 0$, we simply obtain

$$\chi = 0 \tag{24.28}$$

EXERCISES

1. Write the general solutions of (24.17) for a monatomic and a diatomic gas.

Answer.

$$\chi_1 = \frac{d}{dH}\left[\chi_{01}\left(\sqrt{2H} + \frac{v}{\sqrt{3}}\right) + \chi_{02}\left(\sqrt{2H} - \frac{v}{\sqrt{3}}\right)\right]$$

$$= \frac{1}{\sqrt{2H}}\left[\dot{\chi}_{01}\left(\sqrt{2H} + \frac{v}{\sqrt{3}}\right) + \dot{\chi}_{02}\left(\sqrt{2H} - \frac{v}{\sqrt{3}}\right)\right]$$

where $H = c^2/(\gamma - 1) = 3c^2/2$. The other case is

$$\chi_2 = \frac{d}{dH}\,\frac{1}{\sqrt{2H}}\left[\dot{\chi}_{01}\left(\sqrt{2H} + \sqrt{\frac{3}{5}}\,\frac{v}{\sqrt{3}}\right)\right.$$

$$\left. + \dot{\chi}_{02}\left(\sqrt{2H} - \sqrt{\frac{3}{5}}\,\frac{v}{\sqrt{3}}\right)\right]$$

$$= -\frac{1}{2\sqrt{2H^3}}\left[\dot{\chi}_{01}\left(\sqrt{2H}+\frac{v}{\sqrt{5}}\right)+\dot{\chi}_{02}\left(\sqrt{2H}-\frac{v}{\sqrt{5}}\right)\right]$$
$$+\frac{1}{2H}\left[\ddot{\chi}_{01}\left(\sqrt{2H}+\frac{v}{\sqrt{5}}\right)+\ddot{\chi}_{02}\left(\sqrt{2H}-\frac{v}{\sqrt{5}}\right)\right]$$

where $H = 5c^2/2$.

2. Investigate the region of reflection of a simple wave from a wall in a monatomic gas.

Solution. Let the incident wave be travelling from left to right. Its equation is

$$x = (v + c)\, t$$

The arguments of the functions $\dot{\chi}_{01}$, $\dot{\chi}_{02}$ are $v + u$ and $v - u$, which can easily be verified by substituting H in terms of c. From (24.28), at the boundary of the incident wave with the interaction domain *3 — 3'* we have $\dot{\chi} = 0$. This condition can be satisfied only by the function $\dot{\chi}_{01}(u - v)$, the argument of which is equal to a definite constant value (the same as in the case of a simple wave). The function $\dot{\chi}_{02}(u + v)$ at this boundary must be equal to zero at any value of the argument, that is, everywhere. Thus

$$\chi_1 = \frac{1}{\sqrt{2H}}\,\dot{\chi}_{01}\left(\sqrt{2H}-\frac{v}{\sqrt{3}}\right)$$

On the wall condition (24.26) is satisfied:

$$x_0 = -\left(\frac{\partial \chi_1}{\partial v}\right)_{v=0} = \frac{1}{\sqrt{6H}}\,\ddot{\chi}_0\left(\sqrt{2H}\right)$$

whence

$$\dot{\chi}_0\sqrt{2H} = x_0\sqrt{3}\int \sqrt{2H}\, d\sqrt{2H} = x_0\frac{\sqrt{3}}{2}\left(\sqrt{2H}\right)^2 + \text{constant}$$

Therefore

$$\dot{\chi}_0\left(\sqrt{2H}-\frac{v}{\sqrt{3}}\right) = x_0\frac{\sqrt{3}}{2}\left(\sqrt{2H}-\frac{v}{\sqrt{3}}\right)^2 + \text{constant}$$

Going over to u, we write

$$\chi_1 = \frac{x_0}{2u\sqrt{3}}\left[(u-v)^2 + \text{constant}\right]$$

In a self-similar simple wave $u - v = u_0$. Hence at the boundary with a simple wave $u - v$ is also equal to u_0. But on this line $\chi_1 = 0$. Therefore we finally obtain the following required expression for χ_1:

$$\chi_1 = \frac{x_0}{2u\sqrt{3}}\left[(u-v)^2 - u_0^2\right]$$

From formulas (24.14) and (24.15) we determine x and t, and thereby in implicit form v and H.

3. Determine the simple wave that appears after reflection of a self-similar simple wave from a wall in a monatomic gas.

Solution. In Figure 29, the reflected simple wave lies to the right of the characteristic *3′-4′-5′*, which continues the rectilinear characteristic *1-3′* of the incident self-similar wave. The intersection point *3′* exists at any velocity of the piston v_0, provided $|v_0| < 3c_0$, that is, if the self-similar wave is expanding not into vacuum. This can be shown with the help of the equation of the characteristic *3-3′* (see Exercise 4). Point *3′* is common to both the incident and reflected waves. For the incident wave $v - u = -u_0$, and for the reflected wave $v + u =$ constant. But at point *3′*, the velocity of which is that of the piston[11], $v = -|v_0|$. Therefore $u = u_0 - |v_0|$. We see from this that the constant of the reflected wave is equal to $-|v_0| + u_0 - |v_0| = u_0 - 2|v_0|$.

The equation of the reflected simple wave has the form

$$x = (v - c)\,t + f_-(v)$$

The function $f_-(v)$ is defined by the boundary condition (24.27):

$$f_- = -\frac{d\chi_1}{dv}$$

where the total derivative is taken along the line *3′-4′-5′*. But this line also belongs to the simple wave, where $v + u =$ constant. Consequently

$$\frac{d\chi_1}{dv} = \frac{\partial\chi_1}{\partial v} + \frac{\partial\chi_1}{\partial u}\frac{du}{dv} = \frac{\partial\chi_1}{\partial v} - \frac{\partial\chi_1}{\partial u}$$

$$= x_0\left\{\frac{2(v-u)}{u\sqrt{3}} + \frac{1}{2u^2\sqrt{3}}\left[(u-v)^2 - u_0^2\right]\right\}$$

To obtain $f_-(v)$ we must substitute into this the value of u from the simple wave equation, that is, $u = u_0 - 2|v_0| - v$, and make use of the relationship between χ_1 and f_-.

4. Determine the location of point *3′* in Figure 30.

Solution. Write the equation of characteristic *3-3′*:

$$\frac{dx}{dt} = v - c = \frac{1}{2}\frac{x}{t} - \frac{3}{2}c_0$$

whence

$$x = -3c_0t + 4\,(x_0c_0t)^{1/2}$$

The equation of the rectilinear characteristic *1-3′* has the form

$$x = (v + c)\,t = \left(c_0 - \frac{4}{3}|v_0|\right)t$$

[11] Remember that between the piston and the simple wave lies a region of steady flow (see Exercise 3, Sec. 23).

We determine the time of intersection of the two characteristics from the equation

$$t_{3'}^{1/2}=\frac{(x_0c_0)^{1/2}}{c_0-|v_0|/3}$$

For it to be satisfied the condition

$$v_0 < 3c_0$$

must hold.

25

SHOCK WAVES

It was shown in Section 23 that a simple compression wave turns into a shock wave. In other words, the initial hypothesis concerning continuous isentropic gas flow is no longer applicable. For a long time this was a cause of wonder, until Rankine and Hugoniot independently developed the theory of shock discontinuities, demonstrating that a discontinuity is not contrary to the laws of mechanics.

Conditions at a Shock Front. The laws of conservation of mass, momentum and energy are sufficient to determine the conditions at a shock front in terms of fluid mechanics. It is convenient to adopt a frame of reference with respect to which the shock front, that is, the discontinuity, is, at a given time, at rest. For example, in steady flow in a Laval nozzle the shock front, if it develops, is at rest with respect to the nozzle (the laboratory reference frame, in terms of [Sec. 5]). We shall restrict ourselves to the case of a gas flowing into a shock front perpendicular to its surface. If several discontinuity surfaces do not intersect at the given point, as sometimes happens in a Laval nozzle, a reference frame can be chosen such that the flow velocity is perpendicular to the plane of the front.

We shall denote by D the velocity of propagation of the front with respect to a stationary reference frame in which the undisturbed gas ahead of the front is at rest. Then, in a reference frame in which the front is stationary the velocity of the gas is equal to $-D$. If v is the velocity of the gas behind the front, again in a stationary reference frame, its velocity relative to the front is equal to $v - D$. We shall denote the quantities referring to the state of the gas ahead of the front by the subscript 0, for example, ρ_0, p_0, c_0, leaving quantities behind the front with no subscript, that is, ρ, p, c. In the frame connected with the front the flow is steady.

The flow of gas into the front is equal to $-\rho_0 D$, and the outflow from the front is $\rho(v-D)$. From this we obtain the first conservation law at the front:

$$\rho(v-D) = -\rho_0 D \tag{25.1}$$

From Eq. (15.15) we can write the expression for the conservation of momentum:

$$p+\rho(v-D)^2 = p_0+\rho_0 D^2 \tag{25.2}$$

Here we used the fact that the velocity has only a component perpendicular to the front.

And, finally, from the law of conservation of energy (15.3),

$$H+\frac{(v-D)^2}{2} = H_0+\frac{D^2}{2} \tag{25.3}$$

Here the expressions for energy per unit of mass flux have been written, taking into account that, from (25.1), the flow of mass across the front is conserved. As in the problem on nonisentropic gas flow through a long pipe (Sec. 20), the energy balance rather than Bernoulli's theorem, similar in form, has been used. As will be shown later, a shock wave corresponds to a nonisentropic process, and Bernoulli's theorem does not hold here.

Velocity of the Front and Velocity of the Gas. From Eq. (25.1),

$$\frac{\rho}{\rho_0} \equiv \frac{V_0}{V} = \frac{D}{D-v} \tag{25.4}$$

where V and V_0 are the respective specific volumes.

Let us now make use of the law of conservation of momentum (25.2). Replacing $(v-D)^2$ according to (25.4), and solving the equation with respect to D, we obtain

$$D = \left(\frac{p-p_0}{\rho_0-\rho_0^2/\rho}\right)^{1/2} = V_0\left(\frac{p-p_0}{V_0-V}\right)^{1/2} \tag{25.5}$$

This is the formula of the velocity of the front with respect to the undisturbed gas. With the help of (25.4) it is then easy to obtain the velocity $D-v$ of the gas with respect to the front:

$$D-v = \frac{V}{V_0}D = V\left(\frac{p-p_0}{V_0-V}\right)^{1/2} \tag{25.6}$$

It is equally simple to determine the velocity of the compressed gas relative to the uncompressed gas, that is, in a fixed reference frame:

$$v = D-(D-v) = [(p-p_0)(V_0-V)]^{1/2} \tag{25.7}$$

Equations (25.5), (25.6), and (25.7) refer to the propagation of a front in any medium. They do not involve the equation of state.

Note that the propagation velocity of a front in a gas and the velocity of the gas itself are quite different, as can be seen from Eqs. (25.5) and (25.7). The shock wave moves relative to the gas.

The Hugoniot Adiabat. The real properties of the medium are contained in Eq. (25.7), which should include the dependence of the enthalpy on the density and pressure of the gas (or, generally, the medium). In an ideal gas

$$H = \frac{\gamma}{\gamma-1}\frac{p}{\rho} = \frac{\gamma p V}{\gamma-1} \tag{25.8}$$

If the expressions (25.5) and (25.6) are substituted into Eq. (25.3), then the law of conservation of energy in a shock front will be expressed only in terms of the thermodynamic quantities p, V and p_0, V_0. The velocity of the wave or gas is not involved.

The curve that describes the dependence of p upon V and the initial parameters of state p_0, V_0 is called the *Hugoniot adiabat* (in the general case that is what any curve joining two quantities on a shock front is called). The meaning of this curve in the p,V-plane is substantially different from the isentrope pV^γ = constant or the isotherm pV = constant. In isentropic compression the gas state actually changes along the isentrope, and the change is reversible. The same is true of the isotherm. In shock compression, the curve $p(V;\, p_0, V_0)$ shows the pressure needed to compress the gas from volume V_0 and pressure p_0 to volume V. But the process of shock compression itself does not follow an adiabat. A Hugoniot adiabat is a locus all points of which are attainable from a given state p_0, V_0 through compression in a shock wave. As mentioned before, the curve can be constructed not only on the p, V-plane but, for example, on the p,v-plane. Often the isentrope is simply called an adiabat. This is not done here so as to avoid a confusion of terms.

Weak Shock Waves. Every sufficiently small disturbance in the state of a gas propagates with the speed of sound relative to the gas. In the limit the velocity of a shock wave also becomes the speed of sound, because $p - p_0$ is replaced by dp, and $(V_0 - V)/V_0^2$ by $d\,(1/V) = d\rho$. Therefore Eq. (25.5) for the velocity of propagation of the wave reduces to (16.7).

An acoustic disturbance takes place isentropically, that is, the derivative $dp/d\rho$ must be calculated at constant entropy. Consequently a chock wave of sufficiently small amplitude propagates through a gas without altering its entropy.

Let us take a segment on the Hugoniot adiabat close to the initial state, that is, to the state ahead of the front (Figure 31). The ratio $(p - p_0)/(V_0 - V)$ is equal to the tangent of the inclination angle of chord *0-1* to the V axis. For weak waves, it transforms in the limit

into the tangent of the inclination angle of the tangent line to the adiabat at point *0*.

But it was just shown that a small segment of the adiabat close to the initial state coincides with the isentrope. It follows from this that the adiabat and the isentrope have a common tangent at the initial point *0*.

Let us determine the order of magnitude of the change in entropy of a relatively small difference $p-p_0$ in a weak shock wave. Without going into the mechanism of shock compression, it must never-

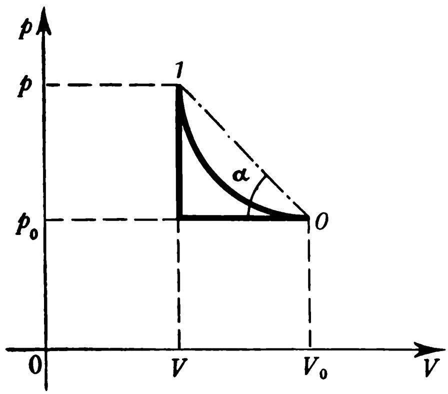


Figure 31

theless be assumed that compression does not occur instantaneously, but along a certain small segment on which the quantities p, V, v vary from the initial to the final state. This assumption, at least, will be made for a weak shock wave. Assuming that a shock compression in fact takes place gradually, we may apply the conservation law to any intermediate instant in the shock transition process and not just to the initial and final states. In particular, if p and V denote the pressure and volume at such an intermediate point, and p_0 and V_0 continue to denote their initial values, we obtain an equation exactly coinciding with (25.5):

$$p-p_0=\frac{D^2}{V_0^2}(V_0-V) \tag{25.9}$$

where D is the same value of the propagation speed of the wave through the gas as in Eq. (25.5). Obviously, at any other velocity at the intermediate point the wave could not travel steadily without changing its profile.

The relationship (25.9) is represented by a straight line joining points *0* and *1*. Therefore, unlike a Hugoniot adiabat, the line *0-1* is the real line of a weak shock transition. (In strong waves, analy-

ses reveal, there are no grounds for assuming that Pascal's law holds in the transition process. Therefore the quantity p on the curve cannot represent the transitional state.)

Let us now make use of the law of conservation of energy in shock compression. As we did in Eq. (25.3), let us substitute for D^2 and $(D - v)^2$ their expressions (25.5) and (25.6) to get

$$E + pV + \frac{V^2}{2}\frac{p-p_0}{V_0-V} = E_0 + p_0V_0 + \frac{V_0^2}{2}\frac{p-p_0}{V_0-V}$$

whence

$$E - E_0 = p_0V_0 - pV + \frac{1}{2}(V_0 + V)(p - p_0)$$

$$= \frac{1}{2}(p + p_0)(V_0 - V) \qquad (25.10)$$

The expression on the right is the area of a trapezoid for which the ordinates of points *0* and *1* are the bases and $V_0 - V$ is the height. A small section of the adiabat *0-1* coincides, as was shown, with the isentrope; hence, from (8.10), the area below the curve is the work done in shock compression. The difference between the energy change and work is, from the thermodynamic relationship (8.16), equal to the temperature T times the change in entropy $S - S_0$. Since temperature is multiplied by a small quantity we need not specify the state it corresponds to in the shock transition.

The difference between the area of the trapezoid and the area of the curvilinear figure is equal to the area of a segment. Its base is a quantity of the first order with respect to the amplitude of the wave $p - p_0$. Then the height of the segment, as is known from geometry, represents a second-order quantity. Consequently the area of the segment, or the change in entropy, is of the third order. The law according to which the quantities change in a weak shock wave closely approximates an isentropic law, that is, the change in a simple wave.

In practice we find that even when the relative amplitude of the shock wave is not too small, that is, when $(p - p_0)/p_0 \sim 1$, the deviations from an isentropic law still has little effect.

The dependence of $S - S_0$ on $p - p_0$ involves an odd (third) power. Therefore the pressure changes in the same direction as the entropy. But according to the second law of thermodynamics entropy increases; hence in shock transitions matter only compresses. Shock waves are always compression waves and never rarefaction waves: only the section of the adiabat lying above the point p_0, V_0 has meaning.

This conclusion is based not only on the second law of thermodynamics but also on the fact that the adiabat passes below its chord, that is, is concave up, which is true of all gases and, in general, the vast majority of substances.

Stability of Shock Waves. We shall now show that only compression shock waves can be stable and travel through a medium without dissipating. The only assumption that must be made in the proof is that the adiabat is concave up, that is, that the curve lies below the chord.

Let us consider the adiabat of an arbitrary shock transition, no longer treated as weak (Figure 32). On the initial section it coincides with the isentrope, having a common tangent. At point *1* the adiabat

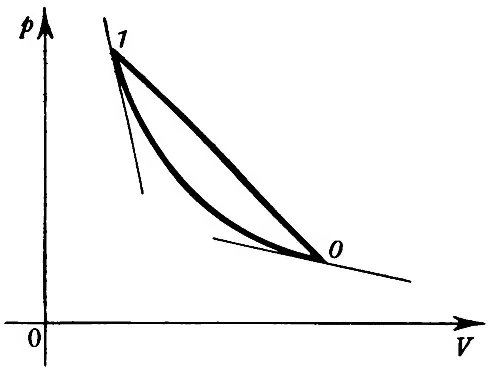


Figure 32

touches another isentrope referring to the shock-compressed substance. It can be seen from the diagram that chord *0-1* is steeper than the lower tangent and less steep than the upper tangent, since it is concave up.

Let us take the corresponding inequality close to the initial point *0*. We write it as follows:

$$\frac{p-p_0}{V_0-V} > -\left(\frac{\partial p}{\partial V}\right)_{S,\,0}$$

Multiply both sides by V_0^2. Taking into account that close to point *0* the relationship $dV/V_0^2 = -d\rho$ holds, and making use of (25.5) and (25.7), we rewrite the inequality as follows:

$$D^2 > \left(\frac{\partial p}{\partial \rho}\right)_{S,\,0} = c_0^2 \tag{25.11}$$

It follows from this that a shock wave travels faster than sound relative to the uncompressed substance.

Near point *1* the initial inequality is reversed:

$$\frac{p-p_0}{V_0-V} < -\left(\frac{\partial p}{\partial V}\right)_{S,\,1}$$

Multiplying by V^2 and replacing dV/V^2 by $-d\rho$, we obtain with the help of (25.6)

$$(D - v)^2 < c^2 \tag{25.12}$$

Hence, a shock wave travels slower relative to the compressed substance than sound disturbances in the same substance.

It follows from the obtained inequalities that a shock wave cannot emit acoustic disturbances ahead of itself. On the contrary, any disturbance following the wave overtakes it. Thus, the characteristics in a compression wave reach the shock front and augment it.

If we imagine a rarefaction shock front or a transition from state *1* to state *0*, the above reasoning is reversed: it will emit acoustic waves ahead of itself and dissipate in space, losing its discontinuity properties, while the acoustic, or simple, waves following it will not catch up and, consequently, will not augment it.

We have thus proved that if the Hugoniot adiabat is concave up, only compression shock waves exist; there are no rarefaction shock waves. In examining a weak wave according to such an adiabat we obtained the same result from the second law of thermodynamics. Now we have the same result from the condition of the stability of a shock front.

Shock compression at a great wave amplitude is an important example of an irreversible process. We encountered an irreversible process in hydrodynamics when speaking of viscosity. But that was a weakly irreversible process, since the velocity gradient was assumed small and the relationship between it and the viscous stress tensor (17.1) was linear. Another example of a weakly irreversible process is heat conductivity at a small temperature gradient. In weak shock waves both these irreversible processes accompany the propagation of the front and determine its structure. The concepts of viscosity and heat conductivity cannot be employed in dealing with the front of a strong wave. The whole transition, as revealed by detailed investigation, which will not be undertaken here, takes place over the free path of a molecule (in a gas). Since the pressure and density change in the process considerably, the gradients cannot be considered small. As for weak waves, they are extended in the ratio $p_0/(p - p_0)$ to the length of the path. We shall accept this statement without proof.

Flow Involving Shock Waves. In Sections 22 and 23 we showed how to solve problems dealing with smooth one-dimensional nonsteady flow. If a flow involves a shock wave the condition of isentropy set in Section 22 disappears: on a shock front entropy is generated. But entropy so produced in a volume of fluid is subsequently transferred together with it, that is, it travels with the same velocity as the fluid itself. This makes it possible to determine the pressure p

in such a body as a function of the density and entropy obtained in the shock transition.

As for the shock front itself, it is given by the differential equation $D = dx/dt = V_0\,[(p - p_0)/(V_0 - V)]^{1/2}$. If this equation has been integrated by numerical methods up to a certain point on the x,t-plane, the next step is to determine D. After that we construct another small line segment representing the path of the shock front. The discontinuities on the line are determined from the formulas obtained in this section. For example, the velocity discontinuity is given by Eq. (25.7). In this way the flow is in principle constructed uniquely.

Shock Waves in a Gas with Constant γ. The equation of the Hugoniot adiabat is especially simple in the case of an ideal gas with a

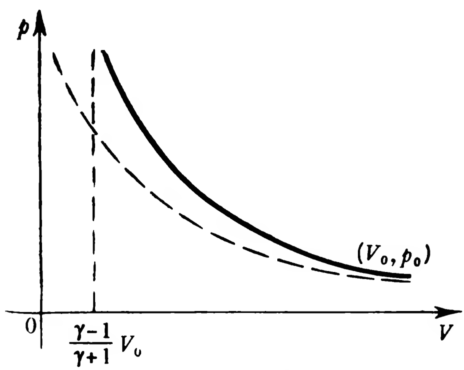


Figure 33

constant specific heat ratio. Using (25.5), (25.6), and (25.8), we obtain

$$\gamma\frac{pV}{\gamma-1}+\frac{V^2}{2}\frac{p-p_0}{V_0-V}=\gamma\frac{p_0V_0}{\gamma-1}+V_0^2\frac{p-p_0}{V_0-V}$$

From this, simple transformations yield either of two relationships:

$$\frac{p}{p_0}=\frac{(\gamma-1)\,V-(\gamma+1)\,V_0}{(\gamma-1)\,V_0-(\gamma+1)\,V} \tag{25.13}$$

$$\frac{V_0}{V}=\frac{(\gamma+1)\,p+(\gamma-1)\,p_0}{(\gamma-1)\,p+(\gamma+1)\,p_0} \tag{25.14}$$

This is the equation of the Hugoniot adiabat in the p,V-plane. It can be seen that at an infinitely great ratio p/p_0 the ratio of the volumes of compressed and uncompressed matter tends to a finite limit

$$\left(\frac{V_0}{V}\right)_{\lim}=\frac{\gamma+1}{\gamma-1} \tag{25.15}$$

For air, for example, this ratio is equal to 6. It would, of course, be true only if the adiabatic exponent really remained constant in the compression. In actual fact, owing to the excitation of vibrational degrees of freedom of the molecules, and also dissociation and ionization, in very strong shock waves air undergoes a ten- or eleven-fold compression. But in any case shock compression has a density limit, shown in Figure 33 as the vertical asymptote to the Hugoniot adiabat. Unlike it, the isentrope asymptotically tends to $V = 0$. Therefore the adiabat is always steeper because in a strong shock wave a large part of the energy is dissipated not on compression but on heating. Such dissipation is irreversible.

EXERCISES

1. Obtain the expression for the change in entropy in a weak shock wave in general form and, in particular, for a gas with a constant adiabatic exponent.

Solution. We proceed from Eq. (25.10):

$$E-E_0=-\frac{1}{2}(p+p_0)(V-V_0)$$

Expand the energy into a series to an accuracy of the term linear with respect to the change in entropy and cubic with respect to the change in volume. Since the multiplier of p is the difference $V - V_0$, the expansion over the volume need be carried out only up to the quadratic term:

$$E-E_0=\left(\frac{\partial E}{\partial S}\right)_V(S-S_0)+\left(\frac{\partial E}{\partial V}\right)_S(V-V_0)$$
$$+\frac{1}{2}\left(\frac{\partial^2 E}{\partial V^2}\right)_S(V-V_0)^2+\frac{1}{6}\left(\frac{\partial^3 E}{\partial V^3}\right)_S(V-V_0)^3$$

$$p-p_0=\left(\frac{\partial p}{\partial V}\right)_S(V-V_0)+\frac{1}{2}\left(\frac{\partial^2 p}{\partial V^2}\right)_S(V-V_0)^2$$

Here the derivatives refer to the initial state. Since

$$\left(\frac{\partial E}{\partial S}\right)_V=T_0, \qquad \left(\frac{\partial E}{\partial V}\right)_S=-p_0$$

after substituting into the initial equation we obtain

$$T_0(S-S_0)=\frac{1}{12}\left(\frac{\partial^2 p}{\partial V^2}\right)_S(V_0-V)^3$$

Thus, for the shock transition to refer to compression the condition $(\partial^2 p/\partial V^2)_S > 0$ must be satisfied. This means that the Hugoniot adiabat must be concave up. From the equation of an isentrope process we have

$$p=p_0V_0^{\gamma}/V^{\gamma}$$

or

$$\left(\frac{\partial p}{\partial V}\right)_S=\gamma(\gamma+1)\frac{p_0V_0^\gamma}{V^{\gamma+2}}$$

Therefore

$$\left(\frac{\partial p}{\partial V}\right)_{S,\ V=V_0}=\gamma(\gamma+1)\frac{p_0}{V_0^2}$$

whence it follows that the change in entropy per unit mass is

$$S-S_0=\frac{p_0\gamma(\gamma+1)}{12T_0V^2}(V_0-V)^3=\frac{N\gamma(\gamma+1)}{12}\,\frac{(V_0-V)^3}{V_0^3}$$

2. Find the change in entropy per unit mass in a shock wave of arbitrary amplitude in a gas with constant γ.

3. Derive the relationship $D(D-v)=v_*^2$ (discovered by Ludwig Prandtl), where v_* is the critical velocity, determined by formula (21.18).

Solution. Introducing the notation $\nu^2\equiv(\gamma-1)/(\gamma+1)$, or $\gamma\equiv$ $\equiv(1-\nu^2)/(1+\nu^2)$, we find that from (21.18)

$$v_*^2=\nu^2D^2+(1-\nu^2)c_0^2$$

Equation (24.14) is written in terms of ν^2 as follows:

$$\frac{\rho_0}{\rho}=\frac{\nu^2p+p_0}{p+\nu^2p_0}$$

whence we find D^2:

$$D^2=\frac{p-p_0}{\rho_0(1-\rho_0/\rho)}=\frac{p+\nu^2p_0}{\rho_0(1-\nu^2)}=\frac{c_0^2}{\nu^2+1}\left(\frac{p}{p_0}+\nu^2\right)$$

With the help of D^2 we determine the critical velocity:

$$v_*^2=\frac{\nu^2}{1+\nu^2}\left(\frac{p}{p_0}+\nu^2\right)c_0^2+c_0^2(1-\nu^2)=\frac{c_0^2}{1+\nu^2}\left(1+\frac{p\nu^2}{p_0}\right)$$

On the other hand,

$$D(D-v)=\frac{p-p_0}{\rho_0(\rho/\rho_0-1)}=\frac{p_0}{\rho_0}\,\frac{(1+\nu^2p/p_0)}{1-\nu^2}=\frac{c_0^2}{1+\nu^2}\left(1+\frac{p\nu^2}{p_0}\right)$$

which proves the required equation.

The critical velocity was determined for the case of isentropic flow. Its appearance in the theory of shock waves is due to the fact that the form of the energy conservation condition on a shock front resembles Bernoulli's theorem and that the Hugoniot adiabat (25.13) involves ν^2.

4. A shock wave in an ideal gas with γ constant strikes a solid, flat obstacle parallel to the shock front. Determine the ratio of the pressure p_1 in the reflected shock wave to the pressure p in the incident wave if in the undisturbed gas the pressure was p_0.

Solution. The velocity of the gas in the incident wave is, from (25.7),

$$v=[(p-p_0)(V_0-V)]^{1/2}$$

After the reflected shock front forms, the gas must come to a halt at the obstacle. Consequently the velocity in the reflected front suffers a disconti-

nuity from v to 0, that is, of the same magnitude as in the incident wave but in the reverse direction. From the same formula (25.7) the discontinuity is expressed in terms of the pressure p_1 and volume V_1 at the reflected shock front as follows:

$$v = -[(p_1 - p)(V - V_1)]^{1/2}$$

Equating the absolute values of both expressions for v, and carrying out simple transformations, we obtain

$$\left(1 - \frac{p_0}{p}\right)\left(\frac{V_0}{V} - 1\right) = \left(\frac{p_1}{p} - 1\right)\left(\frac{V_1}{V} - 1\right)$$

We determine the ratios V_0/V and V_1/V from the equations of the preceding problem and substitute them into the obtained equation:

$$\frac{(p_0/p - 1)^2}{p_0/p + \nu^2} = \frac{(p_1/p - 1)^2}{p_1/p + \nu^2}$$

Next we find the common denominator and transfer all the terms to one side of the equation to obtain

$$\frac{p_1 - p_0}{p}\left[\left(\frac{p_1}{p} - 1\right)\left(\frac{p_0}{p} - 1\right) + (1 + \nu^2)\left(\frac{p_1 + p_0}{p} - 2\right)\right] = 0$$

Cancelling out $p_1 - p_0$, which is not equal to zero, and solving the equation with respect to p_1/p, we obtain the required expression

$$\frac{p_1}{p} = \frac{1 + 2\nu^2 - \nu^2 p_0/p}{\nu^2 + p_0/p}$$

If the incident wave is so strong that $p_0/p \ll \nu^2$, then

$$\frac{p_1}{p} = \frac{3\gamma - 1}{\gamma - 1}$$

For air this ratio equals 8. Such an increase in pressure accounts for the great destructive force of strong shock waves on obstacles.

If the wave is weak the excess pressure $p - p_0$ in the reflected wave doubles. The same result is obtained for an isentropic acoustic wave.

26

APPLICATIONS OF THE THEORY OF SHOCK WAVES

Shock Tube. Shock waves are studied experimentally with the help of a simple facility the principle of which is shown in Figure 34. A compressed and a rarefied gas are separated by a partition which is abruptly punctured (Figure 34*a*). As a result a rarefaction wave forms

in the compressed gas, and a simple wave in the uncompressed gas. If the ratio of the initial motions was large, the shock wave is strong. The ratio p/p_0 for it is large. Let us show how to calculate the amplitude of the wave.

Figure 34*b* schematically presents the pressure distribution in the gas after the removal of the partition. To the left of point *1* the initial gas is not disturbed by the rarefaction wave. The rarefaction wave is between points *1* and *2*. Between points *2* and *3* the flow is steady. Point *3* refers to those particles through which the initial

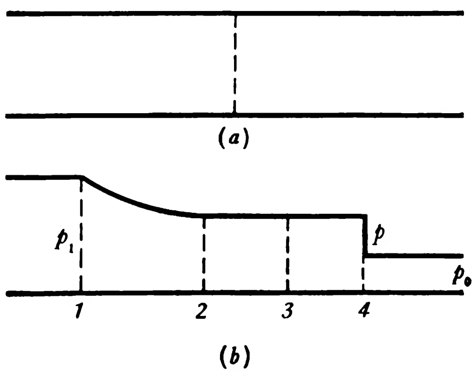


Figure 34

boundary between the gases passed. As in the problem on uniform withdrawal of a piston from a cylinder, a region of steady flow must appear, because the extreme characteristic of the wave does not pass along the boundary of the gas through which the wave propagates. In Exercise 3, Section 23, this characteristic formed an angle with the line of motion of the piston in the x,t-plane; in the present case the angle is with the line of motion of the initial boundary between the gases.

Between points *3* and *4* lies a region of steady flow in the rarefied gas, and the shock wave compressing the gas from its initial state is at point *4*.

At the boundary between the gases (point *3*) their pressures and velocities on both sides coincide. This follows from Newton's Third Law and the condition of continuity of motion. The velocity of the gas on the right is determined from the Hugoniot adiabat:

$$v=[(p-p_0)(V_0-V)]^{1/2}=\left[p_0V_0\left(\frac{p}{p_0}-1\right)\left(\frac{V_0}{V}-1\right)\right]^{1/2} \tag{26.1}$$

The third factor on the right is easily expressed with the help of (25.14), which yields

$$v = c_0 \frac{(2/\gamma)^{1/2} \, (p/p_0 - 1)}{[(\gamma+1)+(\gamma-1)\, p/p_0]^{1/2}} \tag{26.2}$$

The velocities of the gas and the pressure at points *2* and *3* are the same, and they can be found from the simple-wave equation. In this wave, $v + u_1 = u_0 =$ constant. If the gas on the left has a constant adiabatic exponent γ_1, then $u = 2c_1/(\gamma_1 - 1)$. The dependence of the velocity of sound on the pressure is found from the isentrope equation, which holds for a simple wave. But the enthalpy H_1 is proportional to the pressure raised to the power $(\gamma_1 - 1)/\gamma_1$. At the same time the enthalpy is proportional to c_1^2. Consequently

$$\frac{c_1}{c_{10}} = \left(\frac{p}{p_1}\right)^{(\gamma_1 - 1)/2\gamma_1} \tag{26.3}$$

From this the velocity in the simple wave is

$$v = u_0 - u = \frac{2}{\gamma_1 - 1} c_{10} \left[1 - \left(\frac{p}{p_1}\right)^{(\gamma_1 - 1)/2\gamma_1}\right] \tag{26.4}$$

Introducing the notations $p_1/p_0 \equiv \alpha$ and $p/p_0 \equiv x$, we arrive at the equation

$$\frac{2}{\gamma_1 - 1} \frac{c_{10}}{c_0} \left[1 - \left(\frac{x}{\alpha}\right)^{(\gamma_1 - 1)/2\gamma_1}\right] = (x-1) \left(\frac{2}{\gamma\,[(\gamma-1)\,x + \gamma + 1]}\right)^{1/2} \tag{26.5}$$

It is not hard to solve this equation by choosing a suitable value of x. Even when the gas is the same on both sides of the partition, only the pressure and velocity coincide at the boundary. The density, temperature, and entropy suffer a discontinuity, but this does not contradict the laws of mechanics. This kind of discontinuity is called a *contact discontinuity*. It dissipates relatively slowly under the action of transfer processes: diffusion and heat conductivity.

Point Blast. Let us assume that a great amount of energy, E_0, is instantaneously released at some point of a gaseous medium (this approximates what occurs in a nuclear explosion). Let us explain what is meant by "great". A shock wave will travel through the gas from the explosion point. If the energy per unit mass transferred by it is very great in comparison with the initial energy of the gas, and the pressure in the shock wave is many times greater than the initial pressure in the gas, while the mass of the gas involved in the wave is already many times greater than the mass of the device that released the energy, we can assume that the energy was released at a single point.

Suppose, for example, that the released energy is 10^{21} erg, which is approximately the amount of energy released in the explosion of the Hiroshima bomb. Let us investigate the instant when the radius of the shock wave was 100 m. The mass of air in such a volume is about 5000 tons, which is obviously much greater than the mass of the bomb. The energy of the air in that volume was initially 10^{19} erg, which is much less than the energy of the explosion. This, then, corresponds to the definition of a *point blast.* For conventional chemical explosives the stated requirements are not satisfied. The conditions can be approximated by exploding a thin wire by means of a powerful electric pulse.

If a point blast occurs in a medium for which the adiabatic exponent may be assumed constant, the problem of propagation of the shock wave has a complete and exact solution, developed by L.I. Sedov. (Less fully the problem was also investigated independently by K.P. Stanyukovich.)

With the assumptions made, the propagation of a wave from a point blast is determined by two constant dimensional parameters: the energy, E_0, and the initial density of the medium, ρ_0. If the adiabatic exponent, γ, is constant, the density of the gas in the front of a strong wave is connected with ρ_0 by the constant ratio (24.15):

$$\rho = \frac{\gamma+1}{\gamma-1}\rho_0 \tag{26.6}$$

That is why no new dimensional parameters appear in the problem. Since the shock wave is very strong and ionizes the gas (if it is air), we must take the value of γ not for room temperatures, that is, not 7/5, but introduce a certain effective quantity. It is approximately 1.1, which corresponds to the limiting 10- or 11-fold compression, mentioned in the previous section.

The variable quantities r, t and the parameters E_0, ρ_0 can be used to write one and only one dimensionless combination:

$$\zeta = r\left(\frac{\rho_0}{E_0 t^2}\right)^{1/5} \tag{26.7}$$

Thanks to this the problem is self-similar. For two point blasts having different initial energies E_0 and different initial densities ρ_0 the quantities corresponding to the same value of the dimensionless variable ζ are connected by a simple similarity relationship derived from (25.7).

As the wave propagates it remains similar to itself. The points with the same value of ζ travel in space, their states changing according to simple laws. Let us find these laws. First write the initial equations of the problem. These are the Euler equations of motion

$$\frac{\partial v}{\partial t} + v\frac{\partial v}{\partial r} = -\frac{1}{\rho}\frac{\partial p}{\partial r} \tag{26.8}$$

the continuity equation in spherical coordinates

$$\frac{\partial \rho}{\partial t}+\frac{1}{r^2}\frac{\partial}{\partial r}r^2\rho v=0 \tag{26.9}$$

and the equation of conservation of entropy at each point of the gas

$$\frac{dS}{dt}=\frac{\partial S}{\partial t}+v\frac{\partial S}{\partial r}=0 \tag{26.10}$$

This is the entropy the gas receives as a result of compression in the shock wave, after which its state changes isentropically, which is expressed by equation (26.10).

From (20.13) the entropy of the gas is equal to ln (p/ρ^γ). This can be obtained by substituting for temperature its expression from the ideal gas law. The law of the constancy of entropy is conveniently written directly for the ratio p/ρ^γ. Consequently

$$\frac{\partial}{\partial t}\frac{p}{\rho^\gamma}+v\frac{\partial}{\partial r}\frac{p}{\rho^\gamma}=0 \tag{26.11}$$

Note that the amplitude of the shock wave changes as the wave advances, so that the entropy of every part of the gas due to shock compression varies from point to point. The set of three equations, (26.8), (26.9), and (26.11), involves three unknowns and is therefore complete.

To this set we must add the boundary conditions. One is given by Eq. (26.6), the other by Eq. (26.7). Neglecting the initial pressure p_0 as compared with the pressure p (a strong wave!), we obtain the formula for the velocity at the wave front:

$$v=\left[p\left(\frac{1}{\rho_0}-\frac{1}{\rho}\right)\right]^{1/2}=\left[\frac{p}{\rho_0}\left(1-\frac{\gamma-1}{\gamma+1}\right)\right]^{1/2} \tag{26.12}$$

As a consequence of similarity the front must correspond, as noted, to a constant value of ζ, which we shall denote by the symbol ζ_0. (The derivation of ζ_0 will be explained later.) Then from (26.7) we conclude that the equation of the front in the r,t-plane has the form

$$r=\zeta_0\left(\frac{E_0t^2}{\rho_0}\right)^{1/5}$$

In differential form it follows from this that

$$D=\frac{dr}{dt}=\frac{2}{5}\frac{r}{t} \tag{26.13}$$

But then from (25.5) we obtain the following boundary equation for the pressure in the wave front:

$$D=\left(\frac{p}{\rho_0-\rho_0^2/\rho}\right)^{1/2}=\left(\frac{p}{\rho_0}\frac{1}{1-(\gamma-1)/(\gamma+1)}\right)^{1/2}=\frac{2}{5}\frac{r}{t} \tag{26.14}$$

The boundary conditions (26.6), (26.12) and (26.14) are sufficient for solving the problem. It is important that they do not disturb its self-similarity.

A Review of the Solution of the Point-Blast Problem. In order to avoid cumbersome calculations only the main results will be presented below. Condition (26.14) suggests the form in which the pressure should be sought. Namely, we must put

$$p=\frac{8}{25}\frac{\rho_0}{\gamma+1}\frac{r^2}{t^2}p_1(\zeta) \tag{26.15}$$

If we square both sides of (26.14) and substitute the result into (26.15), it is obvious that

$$p_1(\zeta_0)=1 \tag{26.16}$$

From Eq. (26.12) it is easy to surmise that for velocity we should make the substitution

$$v=\frac{4}{5(\gamma+1)}\frac{r}{t}v_1(\zeta) \tag{26.17}$$

Indeed, from (26.15) and (26.17) we obtain that

$$v_1(\zeta_0)=1 \tag{26.18}$$

Finally, it is obvious from (26.6) that ρ must be sought in the form

$$\rho=\frac{\gamma+1}{\gamma-1}\rho_0\rho_1(\zeta) \tag{26.19}$$

where $\rho_1(\zeta_0)=1$.

If we now substitute (26.15), (26.17), and (26.19) into the gas dynamics equations (26.8), (26.9), and (26.11), r and t cancel out separately, just as in the problem of diffusion from an instantaneous point source (Exercise 3, Section 17). There remain three ordinary differential equations involving the independent variable ζ and three dependent variables, p_1, v_1, and ρ_1. But we shall not write these equations explicitly.

It is readily apparent that they involve derivatives with respect to ζ only in the combination $\zeta(d/d\zeta)=d/d\ln\zeta$; ζ is not involved separately. Consequently, if we introduce a new independent variable, $\xi\equiv\zeta/\zeta_0$, the boundary conditions (26.16), (26.18), and the condition for ρ_1 derived from (26.19) will correspond to $\xi=1$.

Therefore the equations can first be integrated without knowing ζ_0, which can then be sought to determine the position of the front. The system of three ordinary differential equations of the first order can be solved with respect to the derivatives $dp_1/d\xi$, $d\rho_1/d\xi$, and $dv_1/d\xi$. If each of the first two equations is divided by the last, the variable ξ is eliminated and two equations remain in which v_1 be-

comes the independent variable, varying from 0 to 1 (in the same way as ξ does).

As is known, the order of a system of equations decreases by one if one of its integrals is known. Such an integral in the posed problem can easily be found from similarity considerations. Take a sphere corresponding to some arbitrary value $\xi < 1$. The sphere contains, all the time, the same portion of the total energy E_0 of the explosion (this is apparent from the similarity provision). If r is the radius of

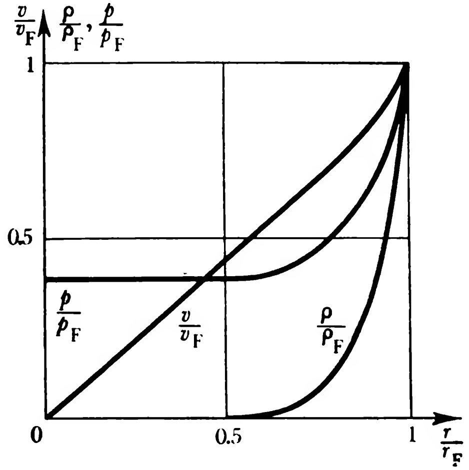


Figure 35

the sphere, the speed of its expansion satisfies Eq. (26.13) with another value of ζ. Such an equation is suitable for any constant ξ. The energy per square centimetre of the surface of a sphere of radius r is therefore equal to

$$\frac{2}{5}\frac{r}{t}\rho\left(\frac{v^2}{2}+E\right)$$

But from (15.13) the energy flux is $\rho v\,(H + v^2/2)$. Expressing E as $p/[(\gamma - 1)\,\rho]$ and H as $\gamma p/[(\gamma - 1)\,\rho]$ and then making use of (26.15), (26.17), and (26.19), we easily obtain, after cancelling out $(r/t)^3$, a nontrivial relationship between the dependent variables p_1, ρ_1, and v_1. It is similar to the energy integral.

After this one differential equation of the first order remains, which can be solved directly in closed form. The solution, however, is extremely cumbersome and we shall not write it out here. Instead, we offer the curves of p/p_F, v/v_F, and ρ/ρ_F as functions of $r/r_F \equiv \xi$ (Figure 35). The subscript F denotes that the quantity is taken at the wave front. It is interesting that p as a function of ξ remains finite

at the centre. But almost all the mass of the gas entrained by the wave front is concentrated in a thin layer close to the front.

The constant ζ_0 is determined as follows. The total energy E_0 of the explosion is

$$E_0 = 4\pi \int_0^{r_F} r^2\,dr \left(\frac{v^2}{2} + \frac{p}{(\gamma-1)\rho}\right) \tag{26.20}$$

Here r_F is the radius of the wave front. Substituting r in terms of the dimensionless variable ξ, and also the expressions p, ρ, and v in terms of p_1, ρ_1, and v_1, we obtain (after cancelling out E_0)

$$4\pi\zeta_0^5 \int_0^1 \frac{8}{25(\gamma+1)^2}\left(v_1^2(\xi) + \frac{p_1(\xi)}{\rho_1(\xi)}\right)\xi^4\,d\xi = 1 \tag{26.21}$$

The integral is simply a number. Hence, from (26.21) we can determine ζ_0.

The time-dependence of the radius of the shock wave, $r \sim t^{2/5}$, is well confirmed by observations.

27

DETONATION WAVES

Detonation. It was shown in Section 25 that shock compression is a strongly irreversible process. The kinetic energy of a substance is dissipated in heating it. A shock wave not sustained by an external energy source eventually damps out, which is the case in explosions.

But if the shock wave propagates through a chemically active substance that reacts to shock compression with the evolution of heat, the situation may change drastically. The heat of the reaction will make up for the losses in the shock wave.

Suppose, for example, that the temperature dependence of the rate of a chemical reaction is subject to the Arrhenius equation, that is, it is proportional to $e^{-A/T}$, where A is the energy of activation of the reaction (see Exercise 2, Sec. 2). The characteristic quantity A has a value of the order of 2 eV, or 23 200 K. Then at room temperature the expression for the rate of the reaction involves the exponential 10^{-33}. This means that in any time interval a molecule is capable of entering into a reaction only during $1/10^{33}$-th of that time. If 1 cm^3 contains 2.7×10^{19} molecules, and each molecule collides with the others 10^9 times per second, then only $2.7 \times 10^{19} \times 10^9 \times$

$\times\ 10^{-33} = 2.7 \times 10^{-5}$ collisions per second, or three collisions in 24 hours, will be effective for the reaction.

When a gas is compressed by a shock wave of sufficient force, its temperature increases. For example, in oxyhydrogen, $2H_2 + O_2$, it reaches approximately 1800 K. Accordingly, the exponential in the Arrhenius equation increases to $10^{-5.5}$, and a molecule reacts in 10^{-4}-10^{-5} s (for a more detailed elaboration of this see Exercise 2). At such a reaction rate the energy is released immediately behind the shock wave and compensates the irreversible losses in shock compression, given a suitable propagation velocity of the wave.

To such a velocity corresponds, of course, a quite definite pressure and compression rate in the wave front or, in other words, a given *detonation regime*.

Here we shall investigate the conditions necessary for a steady-state regime. Then the propagation velocity of the detonation wave can be computed.

It was mentioned in Section 25 that the width of a strong shock front in a gas is of the same order of magnitude as the free path of a molecule, that is, about 10^{-5} cm. But if the reaction takes place in 10^{-5} s, in that time the front advances a distance of the order of 1 cm. Consequently, the reaction takes place not on the shock front itself but in a zone of considerable width behind it. The shock compression only ignites the gas.

Furthermore, if the width of the reaction zone is much greater than the free path, the whole zone can be treated as a continuous medium subject to the equations of gas dynamics, that is, Eqs. (15.6) and (15.11). Only the isentropic equation is now inapplicable since the motion is accompanied by an irreversible chemical reaction. Instead of (15.9) we need an equation for the rate of evolution of energy as a function of the parameters of the gas state, the equation of state, and, of course, the expression for energy in terms of p and V.

But we shall not investigate the course of the reaction in such detail. For conclusions of a general character the steady-state conditions as well as the properties of shock waves obtained in Section 25 are sufficient.

In steady motion the conservation laws must hold for any pair of points in the chemical reaction zone behind the shock front. We already applied this principle in Section 25 in considering a weak shock wave extended in space. But now in the energy balance we must take account of the energy evolved as a result of the chemical reaction.

The reaction, too, must be steady for the whole regime to be steady and the wave to propagate at constant velocity.

Let us construct two Hugoniot adiabats: one for the points located immediately behind the shock compression front, where the chemical reaction has not yet begun, the other for a point, where the reaction

has already ended, that is, where the chemically active substances have "burned out" (Figure 36). The first of these adiabats (*0-1*) in no way differs from the adiabat in Figure 32; the second (*2-3*) is located above it. This follows from the fact that in burning out, the internal energy of molecular motion increases and, consequently, the product pV increases. That is why curve *2-3* lies above curve *0-1*.

Join points *0* and *1* by a straight line, the equation of which coincides with (25.9). This equation is derived solely from the mass and

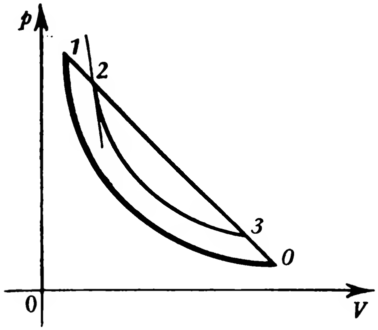


Figure 36

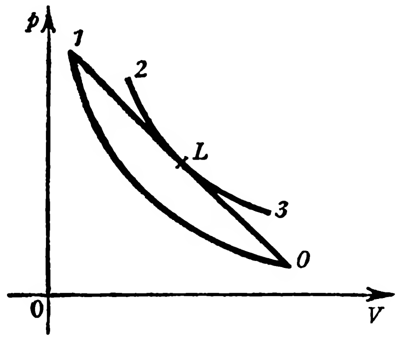


Figure 37

momentum conservation laws, which have the same form at any point of the wave and do not depend on the chemical reaction. As stated in Section 25, the shock transition in a strong wave suffers a discontinuity and does not follow the line *0-1* (such a transition cannot be described in terms of scalar, or Pascal, pressure p). But the subsequent change of state in the reaction zone takes place smoothly and, in steady-state conditions, follows a straight line. The reaction must end at the intersection of the straight line *0-1* with the adiabat *2-3* (towards which the state changes from point *1*) since the chemically active substance is burning out.

It will be proved that, in fact, the straight line *0-1* does not intersect with the adiabat *2-3* but touches it at point L (Figure 37). For this we must examine other possibilities and exclude them.

Suppose the straight line *0-1* intersects the adiabat *2-3* as shown in Figure 36. A comparison with Figure 32 indicates that at the upper intersection point the velocity of the leading front of the shock wave is less than the velocity of sound in the combustion products. But after the reaction is over its products begin to expand. In other words, a rarefaction wave follows the point of total burnout. But such a wave is always unsteady, as we saw in Section 23. Also, a rarefaction wave propagates through matter with the velocity of sound. If this is faster than the propagation of the shock wave, the rarefaction

will overtake it, penetrate the reaction zone, and slow down the reaction. But this would make a steady reaction impossible. Consequently, in a steady-state regime line *0-1* has no upper intersection point with the total burn-out adiabat *2-3*.

Ya. B. Zeldovich suggested a simple reasoning showing why the reaction cannot end at the lower intersection point of line *0-1* and the total burn-out adiabat either. If the process is a smooth one, the states would have to correspond to the points of the line segment lying above the adiabat *2-3*. But for that the evolved energy would have to be greater than the initial total chemical energy. This is seen from the construction in Figure 36. A discontinuous transfer from the upper intersection point to the lower is also impossible, since that would correspond to a rarefaction shock wave, which does not exist if the adiabat is concave up (and in gases this is always the case). But it is nevertheless necessary to reach adiabat *2-3*, since the reaction comes to an end. Thus there remains only one possibility: line *0-1* touches adiabat *2-3*.

The Chapman-Jouguet Condition. We shall now consider the gas-dynamic corollaries of the fact that line *0-1* touches the adiabat *2-3* at point L. On small sections of the adiabat *2-3*, corresponding to total burn-out, no increase in entropy occurs, since the irreversible chemical reaction is completed. Line *0-1* close to the point of osculation coincides with the adiabat up to second-order terms. Since the entropy increases in the reaction, on the straight line it attains its maximum at point L.

The osculation condition of the straight line and the adiabat has the form

$$\frac{p-p_0}{V_0-V}=-\left(\frac{\partial p}{\partial V}\right)_L \tag{27.1}$$

Since Eq. (27.1) satisfies all points on the straight line, let the pressure refer to point L. Multiplying both sides of the equation by V_L^2, we obtain

$$V_L^2\frac{p_L-p_0}{V_0-V_L}=-V_L^2\left(\frac{\partial p}{\partial V}\right)_L=\left(\frac{\partial p}{\partial \rho}\right)_L \tag{27.2}$$

Since the derivative is taken close to the entropy maximum, we should assume that $(\partial p/\partial\rho)_L=(\partial p/\partial\rho)_S$. Therefore Eq. (27.2) involves the square of the velocity of sound, c_L^2, at the *Jouguet point*. A comparison with (25.6) reveals that the left-hand side of the equation is equal to the square of the difference $(D-v_L)^2$, that is, the velocity of the detonation wave relative the combustion products at the burn-out point. Finally

$$D-v_L=c_L \tag{27.3}$$

The velocity of the detonation wave relative to the burn-out products is equal to the velocity of sound in those products. This is the gas-dynamical meaning of the osculation condition, known as the *Chapman-Jouguet condition.*

The unsteady rarefaction wave behind point L cannot as yet overtake the shock wave front, or even simply penetrate the reaction zone. The boundary lies precisely on the Jouguet point.

Condition (27.3) together with the equation of the Hugoniot adiabat are just sufficient to determine the velocity of a steady detonation wave, D (Exercise 3).

Thermonuclear Detonation. In Section 2 we derived the expression for the time rate of a thermonuclear reaction Eq. (2.35). It somewhat resembles the Arrhenius equation. In this connection the question arises whether a nuclear detonation wave is possible. It should be taken into account that at the high temperature which such a wave should generate, radiation accounts for a large part of the energy (see (4.9) and (4.10)). This produces a corresponding lowering of the temperature of the medium containing the radiation, thereby slowing the thermonuclear reaction. At a low reaction rate the reaction zone behind the shock front may extend to an unrealistic degree.

Nevertheless, computations reveal that a mixture of tritium and deuterium would probably be capable of sustaining a thermonuclear reaction in the form of a detonation wave, since the effective cross section of the reaction $H^3 + H^2 = He^4 + n$ is very great. Pure deuterium, it is estimated, should not detonate.

EXERCISES

1. Develop ordinary differential equations for the functions p_1, v_1, an ρ_1 in the problem of a strong explosion. Show that the constructed energy integral satisfies these equations.

2. Determine by how many times the rate of a chemical reaction will increase in a detonation wave propagating at 2.8 km-s^{-1} in a gas with a constant adiabatic exponent 7/5 and an activation energy $A = 2$ eV (neglect excitation of vibrational degrees of freedom). The initial volume of the gas is $V_0 = 2 \times 10^3$ cm^{3}g^{-1}, the initial pressure is $p_0 = 10^9$ dyne-cm^{-2}s^{-1}, and the temperature is 300 K.

Solution. The sound velocity in the initial gas is

$$c_0 = (\gamma p_0 V_0)^{1/2} = 5.3 \times 10^4 \text{ cm-s}^{-1}$$

The pressure in the detonation front is related to the velocity of the wave by the equation developed in Exercise 2, Section 25:

$$D^2 = \frac{c_0^2}{\nu^2+1}\left(\frac{p}{p_0}+\nu^2\right)$$

whence $p/p_0 = 32.5$. The volume of the compressed gas is equal to

$$\frac{V_0}{V} = \frac{\nu^2+p/p_0}{\nu^2 p/p_0+1} \approx 5.1$$

Consequently, $T/T_0 = 6.4$ and $T = 1920$ K.

The exponential in the Arrhenius equation varies from $10^{-31.8}$ to $10^{-3.4}$, so that the reaction accelerates by 28 orders. Compared to this it is of no great consequence that the additional acceleration is achieved by the compression of the gas, which increases the number of collisions among the molecules.

3. Determine the detonation speed in a gas with a constant adiabatic exponent, γ, neglecting the initial energy and pressure of the gas.

Solution. Denote the heat of reaction Q. Since we neglect the initial enthalpy, the energy at the Jouguet point and at the initial state is the same:

$$H_L + \frac{(D-v_L)^2}{2} = Q + \frac{D^2}{2}$$

Now, substituting c_L for $D - v_L$ according to the Jouguet condition and expressing D from Eq. (24.5), we obtain

$$\frac{\gamma p_L V_L}{\gamma-1} + \gamma p_L V_L = Q + \frac{p_L V_0^2}{V_0 - V_L}$$

We again invoke the osculation condition (27.1) and (25.6):

$$(D-v_L)^2 = V_L^2\,\frac{p_L}{V_0-V_L} = V_L^2\,\frac{\gamma p_L}{V_L}$$

From this we express the volume at the Jouguet point and substitute it into the energy equation. We then arrive at the expressions for p_L and V_L:

$$V_L = \frac{\gamma}{1+\gamma}\,V_0, \qquad p_L = \frac{2Q(\gamma-1)}{V_0}$$

Finally, substituting p_L into (25.5), we express the velocity of the detonation wave in terms of the reaction heat:

$$D = V_0\left(\frac{p_L}{V_0-V_L}\right)^{1/2} = [2Q(\gamma^2-1)]^{1/2}$$

Since we have neglected the initial pressure and initial enthalpy of the mixture, the expression obtained does not depend upon the initial density of the mixture either.

PART III

ELECTRODYNAMICS OF CONTINUOUS MEDIA

28

GENERAL EQUATIONS

The study of electrodynamic phenomena in continuous media provided the basis for the discoveries of the elementary laws of electromagnetism. This always required a greater or lesser abstraction from the real properties of matter. Owing to the atomic structure of matter the electric and magnetic fields at every point of a medium vary in a complex and irregular way. Usually, however, it is only the mean values of the fields that are observable in small, but macroscopic, volumes of bodies. The electrodynamics of continuous, or "material", media, unlike the electrodynamics of vacuum, studies the interdependencies between mean quantities.

The operation of averaging in one way or another involves statistical laws, which depend on the structure of the matter in which the electromagnetic processes are taking place.

The electrodynamics of continuous media cannot, therefore, formulate such general laws as the electrodynamics of vacuum. The averaging process performed in this section is in many ways formal and does not lead to a closed set of equations. The relationships obtained can be treated only as points of departure. Their application to concrete conditions and media always require a detailed analysis.

Determination of the Mean. We employed the concept of an average quantity over a volume in Section 15, in the mechanics of a continuous medium. In this connection we must clarify the definition of a macroscopically infinitesimal volume. Such a volume contains a very great number of atoms or molecules and is therefore sufficiently large for us to neglect irregular field fluctuations on the

scale of a single atom. At the same time it is sufficiently small for averaging-out operations over its separate parts—any two halves, for example—to yield the same result. That is why it is called "infinitesimal".

Let us take a cube-shaped volume with side a. The mean value of a certain function f at a point with coordinates x, y, z, and at time t is then

$$\bar{f}(x, y, z, t)=\frac{1}{a^3}\int\int\int_{-a/2}^{a/2} d\xi\, d\eta\, d\zeta\, f(x+\xi, y+\eta, z+\zeta, t) \tag{28.1}$$

Since x is involved in the integral as a parameter, the partial derivative of $\bar{f}$ with respect to coordinate x is equal to the mean value of the derivative of f with respect to x, that is,

$$\frac{\partial \bar{f}}{\partial x}=\overline{\frac{\partial f}{\partial x}} \tag{28.2}$$

It is also obvious that

$$\frac{\partial \bar{f}}{\partial t}=\overline{\frac{\partial f}{\partial t}} \tag{28.3}$$

With the help of the latter two formulas Maxwell's equations [(12.30)-(12.33)] acquire, after averaging, the form

$$\operatorname{curl}\overline{\mathbf{E}}=-\frac{1}{c}\frac{\partial\overline{\mathbf{H}}}{\partial t} \tag{28.4}$$

$$\operatorname{div}\overline{\mathbf{H}}=0 \tag{28.5}$$

$$\operatorname{curl}\overline{\mathbf{H}}=\frac{1}{c}\frac{\partial\overline{\mathbf{E}}}{\partial t}+\frac{4\pi}{c}\overline{\rho\mathbf{v}} \tag{28.6}$$

$$\operatorname{div}\overline{\mathbf{E}}=4\pi\bar{\rho} \tag{28.7}$$

Here we have made use of the fact that, from (28.3) and (28.4), the mean value of the derivative of a certain quantity is equal to the derivative of the mean value of that quantity.

Electric Polarization. Equations (28.4)-(28.7) are conveniently written in other notations that would not include the mean values of the microscopic charge density $\bar{\rho}$ or current density $\overline{\rho\mathbf{v}}$.

For this we must take into account that matter with no charge consists of equal numbers of positive and negative charges. An appreciable excess of charges of any sign would make it mechanically unstable owing to the Coulomb repulsion forces. However strongly an electrified body may be charged, this charge represents but a

negligible part of the charge of the same sign present in the body in the neutral state.

Consider a body that is electrically neutral as a whole.

In an electromagnetic field its own charges may redistribute somewhat over the volume, their mean density at every point becoming other than zero. But since the body as a whole is neutral,

$$\int \bar{\rho}\, dV = 0 \tag{28.8}$$

where the integral is taken over the whole volume of the body.

Let us now introduce the following notation:

$$\bar{\rho} \equiv -\operatorname{div} \mathbf{P} \tag{28.9}$$

where **P** is called the electric polarization of the body. Obviously, the identity (28.9) does not define **P**, since one equation cannot define three vector components.

Vector **P** is conveniently additionally defined as follows. Formula [16.21] shows that the dipole moment of an uncharged body is uniquely defined irrespective of the choice of origin of the coordinate system. Since the charge of a volume element is equal to $\bar{\rho}\, dV$, from [16.17], after going over to a continuous charge distribution, the dipole moment of the body is

$$\mathbf{d} = \int \mathbf{r}\bar{\rho}\, dV \tag{28.10}$$

Let us show that **P** can be expressed in terms of the dipole moment density, that is

$$\mathbf{P} = \bar{\rho}\mathbf{r} \tag{28.11}$$

In Eq. (28.10) we substitute div **P** for $\bar{\rho}$ in accordance with (28.9):

$$\mathbf{d} = -\int \mathbf{r} \operatorname{div} \mathbf{P}\, dV \tag{28.12}$$

Repeating the reasoning employed in deriving the Gauss theorem [11.6], we can transform the space integral involving the operation ∇, which operates on the whole of the integrand, to the integral with respect to the surface $d\mathbf{S}$ confining the volume. The operator ∇ in the space integral is substituted by $d\mathbf{S}$ in the surface integral. Integral (28.12) is then transformed by parts[1]:

$$\mathbf{d} = -\int \mathbf{r}\,(\nabla_{\mathbf{r},\mathbf{P}} \cdot \mathbf{P})\, dV + \int (\mathbf{P}\cdot\nabla)\,\mathbf{r}\, dV$$

$$= -\int \mathbf{r}\,(d\mathbf{S}\cdot\mathbf{P}) + \int (\mathbf{P}\cdot\nabla)\,\mathbf{r}\, dV \tag{28.13}$$

[1] Remember that a subscript of ∇ indicates the quantity on which ∇ operates.

But the surface can be chosen outside the body, in vacuum. Clearly, this will not affect the space integral (28.12), and the integral with respect to the surface becomes zero. From [11.36]

$$(\mathbf{P}\cdot\nabla)\,\mathbf{r} = \mathbf{P}$$

Therefore

$$\mathbf{d} = \int \mathbf{r}\bar{\rho}\,dV = \int \mathbf{P}\,dV \tag{28.14}$$

Thus **P** can be defined as the density of the electric dipole moment.

Magnetic Polarization. The mean current density can be identically expressed in terms of the electric polarization vector and the similar magnetic polarization vector (or the density of the magnetic dipole moment).

From the charge conservation law [12.18], after averaging we obtain

$$\frac{\partial\bar{\rho}}{\partial t} = -\operatorname{div}\overline{\rho\mathbf{v}} \tag{28.15}$$

Making use of the identity (28.9), we rewrite (28.15) as follows:

$$\operatorname{div}\left(-\frac{\partial\mathbf{P}}{\partial t}+\overline{\rho\mathbf{v}}\right)=0 \tag{28.16}$$

This equation is identically satisfied if we put

$$-\frac{\partial\mathbf{P}}{\partial t}+\overline{\rho\mathbf{v}}\equiv c\operatorname{curl}\mathbf{M} \tag{28.17}$$

because $\operatorname{div}\operatorname{curl}\mathbf{M}\equiv 0$.

But (28.17) does not fully define **M**. For **M** it is convenient to take the density of the magnetic dipole moment, **m**, [17.19]:

$$\mathbf{m}=\frac{1}{2c}\int(\mathbf{r}\times\overline{\rho\mathbf{v}})\,dV \tag{28.18}$$

Substitute $\overline{\rho\mathbf{v}}$ from Eq. (28.17) to get

$$\mathbf{m}=-\frac{1}{2c}\int\left(\mathbf{r}\times\frac{\partial\mathbf{P}}{\partial t}\right)dV+\frac{1}{2}\int(\mathbf{r}\times\operatorname{curl}\mathbf{M})\,dV \tag{28.19}$$

We take the partial derivative sign outside the first integral and substitute vector **P** with its expression (28.11). This leaves $\rho\mathbf{r}\times\mathbf{r}\equiv 0$ under the integral sign. We represent the second integral as follows:

$$\int(\mathbf{r}\times\operatorname{curl}\mathbf{M})\,dV=\int\mathbf{r}\times(\nabla_{\mathbf{r},\,\mathbf{M}}\times\mathbf{M})\,dV-\int\mathbf{r}\times(\nabla_{\mathbf{r}}\times\mathbf{M})\,dV$$
$$=\int\mathbf{r}\times(d\mathbf{S}\times\mathbf{M})-\int\mathbf{r}\times(\nabla_{\mathbf{r}}\times\mathbf{M})\,dV$$

Here the product $dv\,\nabla$ is again replaced by $d\mathbf{S}$. Since in the remaining space integral we have a vector product, we expand the product, retaining the order of the factors, remembering however that $\nabla_{\mathbf{r}}$ affects only $\mathbf{r}$:

$$\mathbf{r}\times(\nabla_{\mathbf{r}}\times\mathbf{M})=\mathbf{M}(\nabla_{\mathbf{r}}\cdot\mathbf{r})-(\mathbf{M}\cdot\nabla_{\mathbf{r}})\,\mathbf{r}=\mathbf{M}\operatorname{div}\mathbf{r}-(\mathbf{M}\cdot\nabla)\,\mathbf{r}$$
$$=3\mathbf{M}-\mathbf{M}=2\mathbf{M}$$

We thus arrive at the required equation

$$\mathbf{m}=\int\mathbf{M}\,dV \tag{28.20}$$

The obtained relationships are of the nature of identities. In deducing them we neglected the specific properties of the medium, aside from the fact that it is on the whole neutral.

Maxwell's Equations in Continuous Media. Let us now substitute into the second pair of Maxwell's equations (28.6) and (28.7) the relationships (28.9) and (28.17), so as to eliminate $\overline{\rho}$ and $\overline{\rho\mathbf{v}}$:

$$\operatorname{curl}(\overline{\mathbf{H}}-4\pi\mathbf{M})=\frac{1}{c}\frac{\partial}{\partial t}(\overline{\mathbf{E}}+4\pi\mathbf{P}) \tag{28.21}$$

$$\operatorname{div}(\overline{\mathbf{E}}+4\pi\mathbf{P})=0 \tag{28.22}$$

The equations acquire a more symmetrical form if we introduce the following notation. We shall call the mean electrical field simply the *electrical field in the medium* and no longer use the averaging bar and simply write $\mathbf{E}$.

Introducing the notation

$$\mathbf{E}+4\pi\mathbf{P}\equiv\mathbf{D} \tag{28.23}$$

we see from (28.22) that the *electric displacement*, $\mathbf{D}$, satisfies the equation

$$\operatorname{div}\mathbf{D}=0 \tag{28.24}$$

To be consistent it is convenient to define the mean value of the magnetic field, $\overline{\mathbf{H}}$ as the *magnetic induction in the medium*, $\mathbf{B}$, since then, by analogy with (28.24), we obtain,

$$\operatorname{div}\mathbf{B}=0 \tag{28.25}$$

To preserve symmetry in the notation of electrical and magnetic quantities, the difference

$$\overline{\mathbf{H}}-4\pi\mathbf{M}\equiv\mathbf{B}-4\pi\mathbf{M} \tag{28.26}$$

should be called the *magnetic field in the medium*, **H**. After that the other two Maxwell's equations are written as follows:

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}}{\partial t} \tag{28.27}$$

$$\operatorname{curl} \mathbf{H} = \frac{1}{c}\frac{\partial \mathbf{D}}{\partial t} \tag{28.28}$$

Together with Eqs. (28.24) and (28.25) this yields a set of Maxwell's equations in a continuous medium. Superficially they appear even more symmetrical than the nonaveraged Maxwell's equations for a vacuum with point charges. Actually this symmetry has been achieved at the expense of the uncompleteness of the equations. The set of Maxwell's equations in a medium must be supplemented by relationships that in one way or another take account of the concrete properties of the medium. These dependencies may be quite dissimilar in different media.

Four-dimensional Notation of the Electrodynamic Equations in a Medium. In deriving the equations of electrodynamics from the equations of microscopic theory we made use of only one property of a medium, its electrical neutrality as a whole. This condition is relativistically invariant with respect to the definition of an electric charge (see [14.22]). It is therefore possible to write the equations in four-dimensional form, from which the Lorentz transformations of electromagnetic quantities in a medium can be derived.

Electric field and magnetic induction are essentially the mean values of the electric and magnetic fields in vacuum microscopically defined with due account of the action of real charges within the medium. But then it is clear that to **B** and **E** there corresponds a tensor F_{ik} similar to [14.31]:

$$F_{ik} = \begin{pmatrix} 0 & B_z & -B_y & -iE_x \\ -B_z & 0 & B_x & -iE_y \\ B_y & -B_x & 0 & -iE_z \\ iE_x & iE_y & iE_z & 0 \end{pmatrix}, \tag{28.29}$$

The averaging operation (28.1) involves the ratio dV/a^3 whose numerator and denominator experience the same Lorentz contraction [13.22]. In four-dimensional form, Eqs. (28.25) and (28.27) obtained by averaging the first pair of Maxwell's equations are similar to [15.15]:

$$\frac{\partial F_{ik}}{\partial x_l} + \frac{\partial F_{kl}}{\partial x_i} + \frac{\partial F_{li}}{\partial x_k} = 0 \tag{28.30}$$

where the subscripts i, k, l take on values from 1 to 4, and $x_4 = ict$.

To write equations (28.24) and (28.28) in four-dimensional form we introduce the following tensor:

$$G_{ik} = \begin{pmatrix} 0 & H_z & H_y & -iD_x \\ -H_z & 0 & H_x & -iD_y \\ H_y & -H_x & 0 & -iD_z \\ iD_x & iD_y & iD_z & 0 \end{pmatrix} \tag{28.31}$$

Then both equations combine into one equation

$$\frac{\partial G_{ik}}{\partial x_k} = 0 \tag{28.32}$$

Boudary Conditions for Maxwell's Equations in a Medium. Let us determine the conditions satisfied by fields and inductions at the

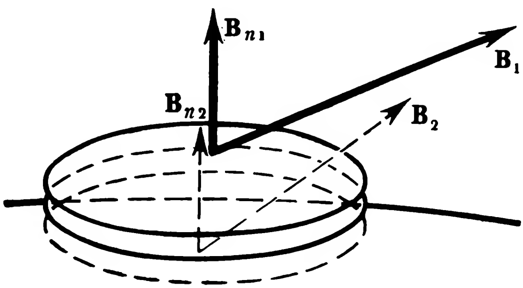


Figure 38

interface between two media. For this we construct a small cylinder such that a section of the interface is within it and parallel to the bases. Integrate Eq. (28.25) over the volume of the cylinder, and make use of the Gauss theorem

$$\int \operatorname{div} \mathbf{B}\, dV = \int \mathbf{B}\, d\mathbf{S}$$

Assuming the height of the cylinder to be of the second order as compared with the radius, we find that the integral over the surface need involve only the bases. Since they are small, the value of vector $\mathbf{B}$ on each of them is constant. We thus obtain (see Figure 38)

$$\mathbf{B}_1\, d\mathbf{S}_1 + \mathbf{B}_2\, d\mathbf{S}_2 = 0$$

The vector $d\mathbf{S}$ is directed along the external normal to the cylinder volume. Therefore

$$d\mathbf{S}_1 = -\, d\mathbf{S}_2 = \mathbf{n}\, dS$$

where $\mathbf{n}$ is a unit normal to the surface dS. Hence

$$(\mathbf{B}_1 - \mathbf{B}_2)\,\mathbf{n} = B_{n1} - B_{n2} = 0 \tag{28.33}$$

Thus at the boundary surface the normal component of the magnetic induction is continuous.

A similar reasoning can be applied to the electric displacement. If there is no outside electrical charge placed on the surface, the normal component D_n is continuous:

$$D_{n1} - D_{n2} = 0 \tag{28.34}$$

But if there is a charge with a surface density σ, then

$$D_{n1} - D_{n2} = 4\pi\sigma \tag{28.35}$$

This follows from the microscopic equation div $\mathbf{E} = 4\pi\rho$ (see [16.1])

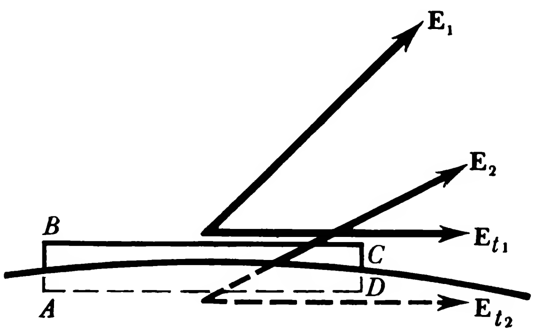


Figure 39

if we involve in ρ only outside charges arbitrarily placed on the interface (for example, electrification by friction).

In order to determine the boundary condition for the electric field, construct a small closed rectangular line enclosing the interface (Figure 39). Integrate equation (28.26) over the surface stretched on the closed line $ABCD$:

$$\int \operatorname{curl}\mathbf{E}\,d\mathbf{S} = -\frac{1}{c}\int \frac{\partial \mathbf{B}}{\partial t}\,d\mathbf{S} = -\frac{1}{c}\frac{\partial}{\partial t}\int \mathbf{B}\,d\mathbf{S}$$

Let the height AB be infinitesimal as compared with the base AD. Then the integral of the finite quantity $\mathbf{B}$ over the area is infinitesimal and can be put equal to zero. The integral of curl $\mathbf{E}$ transforms according to Stokes' theorem [11.17]:

$$\int \operatorname{curl}\,\mathbf{E}\,d\mathbf{S} = \int \mathbf{E}\,d\mathbf{l} = 0 \tag{28.36}$$

which reduces to an integral over the two bases AD and CB for which $d\mathbf{l}_1 = -\,d\mathbf{l}_2$. Since both these segments are small, the vector

for each of them can be taken outside the integral sign:

$$\mathbf{E}_1\, d\mathbf{l}_1 - \mathbf{E}_2\, d\mathbf{l}_2 = (\mathbf{E}_1 - \mathbf{E}_2)\, d\mathbf{l}_1 = 0$$

The closed line $ABCD$ can be turned any way, provided the interface lies between the bases AD and CB. Therefore any component of vector $\mathbf{E}_1 - \mathbf{E}_2$ lying on the interface is zero. Thus we finally have

$$\mathbf{E}_{t1} - \mathbf{E}_{t2} = 0 \tag{28.37}$$

where $\mathbf{E}_{t1}$ and $\mathbf{E}_{t2}$ are two-dimensional vectors tangent to the surface.

After a similar operation with Eqs. (28.28) we obtain

$$\mathbf{H}_{t1} = \mathbf{H}_{t2} \tag{28.38}$$

There may be cases when a current concentrated in a microscopically thin layer flows along the very interface between the media. In that case

$$\mathbf{H}_{t1} - \mathbf{H}_{t2} = \frac{4\pi}{c}\, \mathbf{j}_t \tag{28.39}$$

Here, $\mathbf{j}_t$ has the dimensions of charge flowing in unit time through unit length of the closed line lying on the surface. The factor $4\pi/c$ is of the same origin as in Maxwell's equation (28.6).

EXERCISES

1. Show that a tensor connected with F_{ik} by the relationship $F^*_{ik} = \varepsilon_{iklm} F_{lm}$ (where F_{lm} is a *dual tensor*) satisfies an equation of the form (28.32). The quantity ε_{iklm} is a completely antisymmetrical tensor [Sec. 11].

Solution. From the definition of a dual tensor we have

$$\frac{\partial F^*_{ik}}{\partial x_k} = \varepsilon_{iklm} \frac{\partial F_{lm}}{\partial x_k}$$

The subscripts k, l, and m are all different, according to the properties of a dual tensor. Hence, involved on the right is a cyclic permutation of these subscripts which, according to (28.30), is equal to zero.

2. Write the Lorentz transformation formulas for fields and inductions.

Solution. By analogy with [15.1*b*] and [15.2*b*] we find that all longitudinal components with respect to the relative velocity of the frames of ref

erence are conserved, while the lateral components transform as follows:

$$\begin{array}{l|l}
E'_x = E_x & H'_x = H_x \\
E'_y = \alpha\left(E_y - \frac{V}{c}B_z\right) & H'_y = \alpha\left(H_y + \frac{V}{c}D_z\right) \\
E'_z = \alpha\left(E_z + \frac{V}{c}B_y\right) & H'_z = \alpha\left(H_z - \frac{V}{c}D_y\right) \\
B'_x = B_x & D'_x = D_x \\
B'_y = \alpha\left(B_y + \frac{V}{c}E_z\right) & D'_y = \alpha\left(D_y - \frac{V}{c}H_z\right) \\
B'_z = \alpha\left(B_z - \frac{V}{c}E_y\right) & D'_z = \alpha\left(D_z + \frac{V}{c}H_y\right)
\end{array}$$

where $\alpha = (1 - V^2/c^2)^{-1/2}$ (see [15.1*a*]-[15.2*b*]).

29

ELECTROSTATICS OF CONDUCTORS

Dielectrics and Conductors. According to their electrical properties, all media fall into two classes: (1) dielectrics, in which statistical equilibrium establishes at a constant value of the mean electrical field, and (2) conductors, in which no equilibrium in an electrical field is reached but current flows. In this section we shall examine the conditions of equilibrium of charges on conductors.

Charge Distribution on Conductors. Suppose a conductor is separated from other conductors by a dielectric medium or vacuum, that is, the conductor is isolated, and placed in an external electric field which is constant in time. In equilibrium the charges on the conductor will distribute in such a way that the field inside the conductor will be zero. In other words, within the conductor the field of these charges completely cancels out the action of the external field.

An excess charge of any sign can be placed on an isolated conductor. In this case, too, the field within the conductor becomes zero. Let us show that in all cases the charge density within a conductor is zero.

Indeed, from Eq. (28.7), given a charge density $\bar{\rho}$ other than zero, there would have to be a flux of vector **E** across a surface within which $\bar{\rho} \neq 0$. This applies both to charges belonging to the conducting medium and to charges placed on the conductor from outside.

But the mean field inside a conductor that is in equilibrium is equal to zero, so that $\bar{\rho}$ can have no value other than zero.

Consequently the charge of a conductor can be concentrated only on its surface. In the preceding section the surface-charge density was denoted σ. If the net charge of the conductor is zero, this density will produce in the conductor a nonzero dipole moment (in a charged system the definition of a dipole moment is not unique). But even when the dipole moment is not zero the space-charge polarization is zero, since $\mathbf{P} = \overline{\rho \mathbf{r}}$ and $\bar{\rho} = 0$. Therefore, together with the electric field within the conductor, the electric displacement $\mathbf{D}$ is also zero. It follows from Eq. (28.35) that, since inside the conductor $\mathbf{D} = 0$, the normal component D_n of the electric displacement close to the surface of the conductor is equal to the surface-charge density:

$$D_n = 4\pi\sigma \tag{29.1}$$

Equation (28.37) shows that the tangential component of the electric field on the conductor surface is zero because it becomes zero inside the conductor ($E_t = 0$). If the conductor is in a vacuum (or in air, which is much the same) the field in the surrounding space coincides with the electric displacement, and it can be assumed that Eq. (29.1) involves the normal component of the electric field.

The Potential of a Conductor. In equilibrium, an external field is constant. From (28.4) and (28.7), it satisfies the equations

$$\operatorname{div} \mathbf{E} = 0 \tag{29.2a}$$

$$\operatorname{curl} \mathbf{E} = 0 \tag{29.2b}$$

To satisfy the latter it is sufficient to put

$$\mathbf{E} = -\operatorname{grad} \varphi \tag{29.3}$$

Then the electrostatic potential φ is determined from the Laplace equation

$$\operatorname{div} \operatorname{grad} \varphi \equiv \nabla^2\varphi = 0 \tag{29.4}$$

On the surface of every conductor the potential in equilibrium assumes a constant value, since otherwise there would exist a component of the potential gradient directed along the surface, that is, a tangential component of the electric field, $E_t = -\operatorname{grad}_t \varphi$. But this is impossible. Any point within the conductor possesses a potential of the same value, since otherwise there would be an electric field $\mathbf{E} = -\operatorname{grad} \varphi$ inside the conductor, which is impossible in equilibrium.

From Eq. (29.3) it can be observed that the potential is determined up to a constant term. This term is conventionally chosen so that

the potential of the earth, or that of a grounded conductor, is zero. If the potentials of other, ungrounded, conductors are given, Eq. (29.4) fully defines the potential between the conductors.

The Energy of a System of Conductors. Continuing to assume that our conductors are in vacuum, let us calculate the energy of the field they create. From Eq. [15.24] we know that this energy is equal to

$$E = \frac{1}{8\pi} \int \mathbf{E}^2 \, dV \tag{29.5}$$

where the integrals are taken over the whole volume not occupied by the conductors.

We replace $\mathbf{E}^2$ by $-(\mathbf{E} \cdot \operatorname{grad} \varphi)$ according to Eq. (29.3) and apply Gauss' theorem in the same way as in Eq. (28.13). In other words, after integrating by parts, we replace $dV\,\nabla$ (where ∇ acts to the whole integrand) by $d\mathbf{S}$ to get

$$E = \frac{1}{8\pi} \int \varphi \, (d\mathbf{S} \cdot \mathbf{E}) + \frac{1}{8\pi} \int \varphi \operatorname{div} \mathbf{E} \, dV \tag{29.6}$$

In the space between the conductors that is free of charges $\operatorname{div} \mathbf{E} = 0$ (see [16.1]); hence the space integral in (29.6) vanishes. To calculate the surface integral, we must take into account that on every conductor the potential φ has a constant value. Using the subscript i to number individual unconnected conductors, we obtain

$$E = -\frac{1}{8\pi} \sum_i \varphi_i \int \mathbf{E} \, d\mathbf{S}_i \tag{29.7}$$

The surfaces of the conductors are external with respect to the volume contained between them. Therefore in the integral (29.6) the surface element $d\mathbf{S}$ is directed inside each conductor. Hence, if $d\mathbf{S}_i$ is the external normal to the surface of the ith conductor, then

$$\mathbf{E} d\mathbf{S}_i = -(\mathbf{E} \cdot \mathbf{n}_i) \, dS_i = -E_{ni} \, dS_i \tag{29.8}$$

But from (29.1) it follows that in vacuum $E_{ni} = 4\pi\sigma_i$, and therefore the energy is

$$E = \frac{1}{2} \sum_i \varphi_i \int \sigma_i \, dS_i = \frac{1}{2} \sum_i e_i \varphi_i \tag{29.9}$$

where e_i is the charge of the ith conductor.

Expression (29.9) differs by the factor 1/2 from the energy of charges in vacuum placed in an external field, expressed by the formula $U = \sum e_i \varphi_i$. That is because in the case of a system of charged conductors the field is created by the charges themselves

and is not imposed from outside. Indeed, by virtue of the linearity of the electrodynamic equations, there is a linear dependence between the potentials and the charges. Supposing that the charges e_i' are increasing from zero to their real values e_i in proportion to their values according to the law $e_i' = \lambda e_i$, we see that $de_i' = e_i d\lambda$, and for the varying potentials φ_i' we obtain the same kind of linear dependence, $\varphi_i' = \lambda\varphi_i$. Therefore

$$E = \sum_i \varphi_i e_i \int_0^1 \lambda\, d\lambda = \frac{1}{2} \sum_i \varphi_i e_i$$

Capacitance and Coupling Factor. The general linear relationship between charges and potentials in a system of conductors, has evidently, the form

$$e_i = \sum_k C_{ik}\varphi_k \tag{29.10}$$

where the quantities C_{ik} are called the *coefficients of capacitance*. They have the dimensions of length.

Suppose all conductors but the kth in a given spatial configuration are grounded, that is, $\varphi_{k'} = 0$ if $k' \neq k$. Then the charge on the ith conductor at $i \neq k$ is

$$e_i = C_{ik}\varphi_k \tag{29.11}$$

The sign of the ith charge must be opposite to the sign of the kth charge since the ith conductor is grounded; thus the charge of the same sign as the kth is removed. It will be readily appreciated that this corresponds to a minimum energy, that is, the equilibrium condition. Consequently,

$$C_{ik} < 0 \quad \text{at} \quad i \neq k \tag{29.12}$$

At $i = k$ the coefficient of capacitance is positive.

It will be shown in Exercise 2 that the matrix $[C_{ik}]$ is symmetrical, that is, $C_{ik} = C_{ki}$. This means that at unit potential the kth conductor induces the same charge e_i on the ith conductor as the unit potential on the ith induces on the grounded kth conductor.

The set of equations (29.10) can be solved for the potentials. The matrix of the coefficients expressing the potentials in terms of the charges on the conductors is conventionally called the inverse of the matrix $[C_{ik}]$ and is denoted $[C_{ik}]^{-1}$. We know from algebra that $[C_{ik}]^{-1}$ is $[\mathrm{C}_{ki}/\det [C_{ik}]]$, where C_{ik} is the cofactor of the element C_{ik} in the determinant $\det [C_{ik}]$. From the definition of an inverse there follows a relationship between the direct and inverse matrices:

$$[C_{ik}]^{-1} [C_{kl}] = [\delta_{il}] \tag{29.13}$$

The quantities $C_{ik}^{-1} \equiv C_{ki}/\det[C_{ik}]$ are called the *coupling coefficients* or *factors*. The potentials are expressed in terms of the charges as follows:

$$\varphi_i = \sum_k C_{ik}^{-1} e_k \tag{29.14}$$

If we substitute the relationship (29.14) into the energy expressed by (29.9), we obtain

$$E = \frac{1}{2} \sum_{i,k} C_{ik}^{-1} e_i e_k \tag{29.15}$$

Energy is expressed in terms of the potentials as follows:

$$E = \frac{1}{2} \sum_{i,k} C_{ik} \varphi_i \varphi_k \tag{29.16}$$

Let us now show the relationship between the capacitance C of a capacitor (condenser) and the coefficients of capacitance of its plates. A capacitor's capacitance C relates its charge to the potential difference between the plates:

$$e = C(\varphi_2 - \varphi_1) \tag{29.17}$$

The energy of a capacitor is, as is known, equal to

$$E = \frac{e^2}{2C} \tag{29.18}$$

If the charge on one plate is e, then the charge on the other is $-e$. Consequently, from equation (29.15), and taking into account the fact that $C_{12} = C_{21}$, we obtain

$$E = \frac{1}{2} (C_{11}^{-1} - 2C_{12}^{-1} + C_{22}^{-1}) e^2 \tag{29.19}$$

Comparing this with (29.18), we obtain the expression for capacitance, first in terms of the coupling factors:

$$\frac{1}{C} = (C_{11}^{-1} - 2C_{12}^{-1} + C_{22}^{-1}) \tag{29.20a}$$

Substituting then the coupling factors from the general definition of an inverse matrix, we finally obtain

$$C = \frac{C_{11}C_{22} - C_{12}^2}{C_{22} + 2C_{12} + C_{11}} \tag{29.20b}$$

The Method of Electric Images. In mathematical physics there are many methods of calculating the electrostatic fields of conduc-

tors. Notably, fields depending on two coordinates (plane fields) are determined by the method of functions of a complex variable examined in Section 15.

Let us examine some problems relating to fields in three dimensions.

Let there be a charge e at a distance a from an infinite grounded plane (Figure 40). Determine the electric field.

To solve the problem the following device can be employed. We place an imaginary charge $-e$ at a point opposite the charge

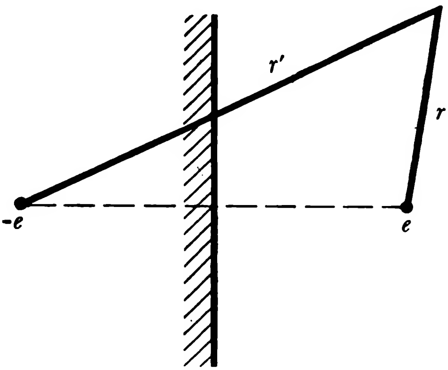


Figure 40

at a distance a behind the conductor (the "image" of charge e in the plane). Then the potential in the half-space in which the real charge is located is equal to

$$\varphi = \frac{e}{r} - \frac{e}{r'} \tag{29.21}$$

where r is the distance from a given point to the real charge, and r' is the distance to its image. By virtue of the symmetrical configuration of the charges, the potential on the conductor is zero, that is, the conductor is really equipotential. The function e/r' satisfies the Laplace equation everywhere in the half-space before the conductor: it has no singular points in that domain. The function e/r also satisfies the Laplace equation everywhere, except the point where the real charge is located. We have thus developed a solution of the Laplace equation that satisfies the boundary condition on the plane.

This problem is a special case of the more general problem of a point charge e in the neighbourhood of a grounded conducting sphere (Figure 41). Here the following construction must be made. Let the distance of the charge from the centre of the sphere be R,

and the radius of the sphere, r_0. Lay off from the centre a line $OA = r_0^2/R$, draw an arbitrary radius OB, and join point B with points A and e. The triangles eOB and AOB are similar since they have a com-

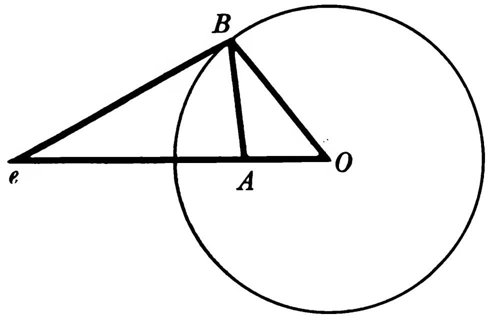


Figure 41

mon apex with the same angle at O, while the adjoining sides of the angle are proportional by construction:

$$\frac{eO}{OB}=\frac{R}{r_0}=\frac{OB}{AO}=\frac{r_0}{AO}$$

Consequently, the third sides are also proportional to them:

$$\frac{Be}{BA}=\frac{R}{r_0}$$

Now at point A place an imaginary charge e' such that

$$e'=-\frac{r_0}{R}e \tag{29.22}$$

Then the potential at an arbitrary point outside the sphere is

$$\varphi=\frac{e}{r}+\frac{e'}{r'} \tag{29.23}$$

where r and r' are the respective distances of the point from the charge and from A.

On the surface of the sphere $\varphi = 0$. To make sure of this it is sufficient to substitute e' from (29.22) into (29.23) and assume $r' = BA$. Then the required condition will be satisfied thanks to the proved proportionality of the sides of the triangles.

If there is a real charge e_1 on the sphere, it is sufficient to add the term e_1/R to the potential (29.23), which does not violate the condition of the constancy of the potential on the sphere.

Nonhomogeneous Conductors. The equipotential condition for a conductor formulated at the beginning of this section holds only

for conductors of homogeneous composition. Let us now consider what is needed for equilibrium in a conductor in which the concentration of current carriers is not homogeneous. In the absence of an electric field the equilibrium condition consists in the requirement that the thermodynamic potential be minimal (Sec. 8). To determine equilibrium volume concentration of certain particles, we must require that the transfer of a small number of these particles dN from one point in space to another does not alter the overall thermodynamic potential G of the system. Then the work done in the transfer, which is equal to the total change in G, is zero:

$$dA=\left[\left(\frac{\partial G}{\partial N}\right)_1-\left(\frac{\partial G}{\partial N}\right)_2\right]dN=0$$

where the subscripts "1" and "2" refer to two arbitrary points. But from (8.48) $\partial G/\partial N = \mu$, where μ is the chemical potential of the transferred particles. Hence, in the absence of an electric field the chemical potential in equilibrium is the same at all points of the system.

If the particles carry a charge and are in an electric field, the work done in their displacement consists in the increment of G and in the change in potential energy:

$$dA=\left[\left(\frac{\partial G}{\partial N}\right)_1-\left(\frac{\partial G}{\partial N}\right)_2+e\varphi_1-e\varphi_2\right]dN=0$$

From this follows the equilibrium condition with respect to the charge carriers:

$$\mu + e\varphi = \text{constant} \tag{29.24}$$

Let, for example, the charge carriers be ions forming a weak solution of variable spatial concentration c. Then the chemical potential μ is determined from Eq. (12.6). The potential difference between points *1* and *2* is

$$\varphi_2-\varphi_1=-\frac{\theta}{e}\ln\frac{c_2}{c_1} \tag{29.25}$$

Contact Potential Difference. Another example of a nonhomogeneous conductor is the case of two contacting metals. The charge carriers in both metals are, of course, identical: electrons.

In order to transfer an electron from one metal to the other, the work done must be equal to the difference between the work A_1 needed to eject the electron from the first metal into vacuum and the work A_2 obtained on the electron's entrance into the other metal. The work A_1 is done partly in the surface layer of the metal and partly by the force of attraction of the charge to its image, when the electron has already escaped from the metal. When the metals are in equilibrium, the total work in the transfer of the electron

must be zero. Therefore, in contacting metals in equilibrium there is a potential difference that compensates the difference between the work functions A_1 and A_2:

$$\varphi_2 - \varphi_1 = \frac{A_2 - A_1}{|e|} \tag{29.26}$$

Here the negative charge of the electron has been taken into account.

The above reasoning is illustrated in Figure 42, which shows the motion of a charge along a closed path. Work along the path is expended in the passage from metal *1* into vacuum, obtained in

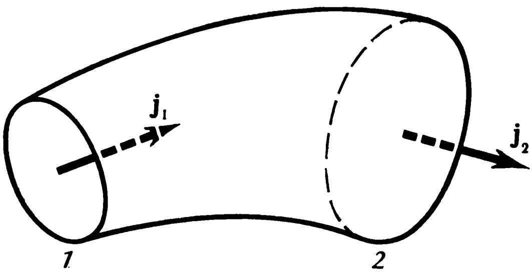


Figure 42

entering metal *2*, and is done along the path *1-2* in the vacuum by the electric field formed between the metals. But in equilibrium the total work over a closed path is zero. If this were not so the system described here would represent a perpetual motion machine: nothing changes in the motion of the electron along the path, so that work would be obtained from nothing, without expending any external energy.

Thus

$$A_2 - |e|(\varphi_2 - \varphi_1) - A_1 = 0$$

which is expressed by Eq. (29.26).

If three metals are in contact one after another, the potential difference between the end metals is

$$\varphi_3 - \varphi_1 = \frac{(A_3 - A_2) + (A_2 - A_1)}{|e|} = \frac{A_3 - A_1}{|e|} \tag{29.27}$$

The work function for the intermediate metal is eliminated from the formula. In particular, if the same metal is on both ends, the potential difference becomes zero, as it should be in accordance with the energy conservation law.

The work function of an electron leaving various faces of a metallic crystal may differ depending on the density of the atoms near the surface. Therefore in vacuum, that is, outside the surface, a po-

tential difference is established between two neighbouring faces with different work functions (but inside a metallic monocrystal is equipotential). In the atmosphere this potential difference is quickly eliminated by ions adhering from outside.

EXERCISES

1. Prove that the minimum or maximum of a potential cannot occur at a point located outside a conductor.

Solution. Let us assume the reverse. Then on a small closed surface surrounding the extremum the potential gradient is everywhere directed either to the outside or to the inside of the surface. The flux of vector $\mathbf{E} = -\operatorname{grad} \varphi$ across the surface is not zero, which is impossible if there is no charge within it. Hence, a charge cannot be in equilibrium in a field of charged conductors if it is not on one of them.

2. Prove that in a system of conductors the differential of the energy change, dE, is equal to $\sum \varphi_i \, de_i$ or $\sum e_i \, d\varphi_i$.

Solution. By definition

$$dE = \frac{1}{4\pi} \int (\mathbf{E} \cdot d\mathbf{E}) \, dV$$

Substitute $-\operatorname{grad} \varphi$ for $\mathbf{E}$ and integrate by parts according to the method described in this section to get

$$dE = -\frac{1}{4\pi} \int \varphi \, d\mathbf{E} \, d\mathbf{S} + \frac{1}{4\pi} \int \varphi \, (\operatorname{div} d\mathbf{E}) \, dV$$

The second integral is zero since in vacuum $d\mathbf{E} = 0$. In the first integral we take advantage of the fact that the potential on the surface of each conductor is constant. This yields

$$dE = -\frac{1}{4\pi} \sum_i \varphi_i \int d\mathbf{E}_i \, d\mathbf{S}_i$$

But by (29.1) $dE_{ni}/(4\pi) = d\sigma_i$, and in the obtained integral the surface element $d\mathbf{S}_i$ is directed inside the conductor. Consequently

$$dE = \sum_i \varphi_i \, d \int \sigma_i \, dS_i = \sum_i \varphi_i \, de_i$$

It is apparent from this that $\varphi_i = \partial E / \partial e_i$. In the initial integral we substitute $-\operatorname{grad} \varphi$ for $\mathbf{E}$ under the differential sign and integrate similarly to get

$$dE = \sum_i e_i \, d\varphi_i$$

Therefore

$$e_i = \frac{\partial E}{d\varphi_i}$$

But $e_i = \sum C_{ik}\varphi_k$, so that

$$C_{ik} = \frac{\partial^2 E}{\partial\varphi_i\,\partial\varphi_k} = C_{ki}$$

3. Determine the coefficients of capacitance and the coupling factors for two concentric spheres of radii r_1 and r_2.

Solution. Denote the charge on the inside sphere e_1 and the charge on the external sphere e'. The potential in the space between the spheres is

$$\varphi = \frac{e_1}{r} + \text{constant}$$

while in the external domain the potential is

$$\varphi = \frac{e'}{r}$$

From the condition that the outer sphere has a unique potential we find that

$$\frac{e_1}{r_2} + \text{constant} = \frac{e'}{r_2}$$

whence

$$\text{constant} = \frac{e'}{r_2} - \frac{e_1}{r_2}$$

Therefore the potential of the inside sphere is

$$\varphi_1 = \frac{e_1}{r_1} + \frac{e'}{r_2} - \frac{e_1}{r_2}$$

and the potential of the external sphere is

$$\varphi_2 = \frac{e'}{r_2}$$

The charge on the inside surface of the external sphere is equal to $-e_1$ because all lines of force from the charge of the inside sphere terminate on it. Consequently, the total charge of the external sphere e_2 equals $e' - e_1$, and hence

$$\varphi_1 = \frac{e_1}{r_1} + \frac{e_2}{r_2}$$

$$\varphi_2 = \frac{e_1}{r_2} + \frac{e_2}{r_2}$$

The coupling factors assume the values

$$C_{11}^{-1} = \frac{1}{r_1}, \qquad C_{12}^{-1} = C_{21}^{-1} = \frac{1}{r_2}, \qquad C_{22}^{-1} = \frac{1}{r_2}$$

The determinant $\det(C^{-1})$ is equal to $r_2^{-1}(r_1^{-1} - r_2^{-1})$. For the coefficients of capacitance we obtain the following values:

$$C_{11} = \frac{r_1 r_2}{r_2 - r_1}, \qquad C_{12} = C_{21} = -\frac{r_1 r_2}{r_2 - r_1}, \qquad C_{22} = \frac{r_2}{r_2 - r_1}$$

4. Given a system of grounded conductors and a point charge e that induces charges e_i' on them. If (1) the charge e is removed, (2) all the conductors except the ith are left grounded, and (3) the potential on the ith conductor is brought up to a certain value φ_i', then at the point where the charge e was formerly located there will be a potential $\varphi_i'(0)$. Express the charge e_i in terms of e, φ_i' and $\varphi_i'(0)$.

Solution. Imagine the point charge as an infinitely small charged conductor whose capacitance can be neglected. From the symmetry of the coefficients of capacitance, $C_{ik} = C_{ki}$, there follows the identity

$$\sum_{i,\,k} C_{ik}\varphi_i\varphi_k' = \sum_i e_i\varphi_i' = \sum_i e_i'\varphi_i$$

In the first case, when all the conductors are grounded, their potentials are zero. Consequently, the point charge was present together with a zero potential:

$$\sum_i \varphi_i e_i' = 0$$

Therefore

$$\sum_i e_i\varphi_i' = e_i\varphi_i' + e\varphi_i'(0) = 0$$

whence we obtain the required expression for e_i:

$$e_i = -e\frac{\varphi_i'(0)}{\varphi_i'}$$

In the case of a spherical conductor this formula turns into (29.22).

5. A point charge is placed between two concentric grounded spheres. Determine the charges it induces on the spheres.

Solution. If the inside sphere is not grounded, the potential between the spheres varies according to the law

$$\varphi = \frac{e_1}{r} + a$$

(we disregard the action of the point charge). Then, assuming the external sphere to be grounded and denoting the radii of the spheres r_1 and r_2, we obtain two equations:

$$\frac{e_1'}{r_1} + a = \varphi_1'$$

$$\frac{e_1'}{r_2} + a = 0$$

whence

$$e_1'\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = \varphi_1', \qquad a = -\frac{e_1'}{r_2}$$

The potential at the place where the point charge e was located is thus equal to

$$\varphi_1'(0)=\frac{e_1'}{r}-\frac{e_1'}{r_2}$$

Hence, the charge induced on the inside sphere when it was grounded was equal to

$$e_1=-e\frac{\varphi_1'(0)}{\varphi_1'}=-e\frac{r^{-1}-r_2^{-1}}{r_1^{-1}-r_2^{-1}}=-e\frac{r_1}{r}\frac{r_2-r}{r_2-r_1}$$

For the charge e_2 on the outer sphere we obtain:

$$e_2=e-e_1=-e\frac{r_1^{-1}-r^{-1}}{r_1^{-1}-r_2^{-1}}=-e\frac{r_2}{r}\frac{r-r_1}{r_2-r_1}$$

For the case of two conducting planes we must replace the ratios r_1/r and r_2/r by unity. The distribution of the induced charge is inversely proportional to the distance to the planes. It would be much more difficult to obtain this result by the method of consecutive images.

6. An infinite charged plane has a semispherical projection. At infinite distance from the projection the surface-charge density on the plane is σ_0. Determine the field and the charge concentrated on the projection.

Solution. Let us turn to the problem on the flow of an ideal fluid around a sphere (Sec. 15). The velocity potential in that problem satisfies the Laplace equation $\nabla^2\varphi=0$. If we pass a median plane through the centre of the sphere parallel to the streamlines at infinity, the lines of force of the field with which we are concerned in the present problem will lie on the equipotential surfaces considered in the problem on the flow around a sphere. To demonstrate this, we determine the electric field by analogy with the velocity field in the form

$$\mathbf{E}=4\pi\sigma_0\mathbf{n}+\frac{3(\mathbf{d}\cdot\mathbf{r})\,\mathbf{r}-r^2\mathbf{d}}{r^5}$$

Where $\mathbf{n}$ is a unit vector normal to the plane, and $\mathbf{d}$ is a vector parallel to it. If the radius of the projection is r_0, at $r>r_0$ the boundary condition $\mathbf{E}_t=0$ must be satisfied on the plane. This follows from the fact that on it $(\mathbf{d}\cdot\mathbf{r})=0$, and $\mathbf{n}$ and $\mathbf{d}$ are perpendicular to it. At $r<r_0$, the same condition refers to the projection. Therefore

$$\mathbf{d}=4\pi r_0^3\sigma_0\mathbf{n}$$

Expressing from this the surface-charge density in terms of $\mathbf{E}$ and integrating over the surface of the projection, we find that the charge on it is equal to $3\pi r_0^2\sigma_0$.

7. Determine the electric field in vacuum between two intersecting faces of a monocrystal for which the difference between the work functions is not zero. The angle between the faces is ψ_0. The length of the edge along which the faces intersect is assumed to be infinite, and both semiplanes are also infinite.

Solution. The potential satisfies the Laplace equation $\nabla^2\varphi = 0$ and acquires constant values on each of the faces. Therefore φ should be sought at any point as a function only of the angle ψ between the plane through that point and one of the faces. The Laplace equation for this case reduces to the form

$$\frac{d^2\varphi}{d\psi^2}=0$$

Therefore

$$\varphi = C_1\psi + C_2$$

We find the integration constants from equation (29.26). Finally we obtain

$$\varphi=\frac{A_2-A_1}{|e|}\frac{\psi}{\psi_0}+\varphi_1$$

The electric field is perpendicular to the plane ψ = constant and is equal to

$$E_\psi=-\frac{1}{r}\frac{d\varphi}{d\psi}=\frac{A_2-A_1}{|e|\psi_0}\frac{1}{r}$$

Close to the edge it is very strong.

30

ELECTROSTATICS OF DIELECTRICS

General Equations. In dielectrics, or insulators, charges come into equilibrium when the mean electric field **E** is not zero (as agreed, we do not write the averaging bar). It follows from this that, according to (28.26), in a constant field

$$\text{curl } \mathbf{E} = 0 \tag{30.1}$$

Hence in the electrostatics of dielectrics we can use the electrostatic potential, φ, defined as

$$\mathbf{E} = -\text{grad } \varphi \tag{30.2}$$

Furthermore, in the absence of extraneous charges in a medium which is on the whole neutral,

$$\text{div } \mathbf{D} = 0 \tag{30.3}$$

The electric displacement **D**, the field **E**, and the electric polarization **P** are connected by the relationship (28.23):

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} \tag{30.4}$$

This set of equations has a solution only when the connection between the polarization (or the displacement) and the field **E** has been established.

Free Energy of a Dielectric. It will be appreciated that the relationship described above must depend on the nature of the dielectric and the external conditions (temperature, pressure). The problem is conveniently approached from the general thermodynamic point of view.

Suppose there is a conductor with a potential φ in a dielectric medium. Considering its charge as an external parameter with respect to the dielectric, we find that the work of changing the charge of the conductor by de is

$$dA = \varphi \, de \tag{30.5}$$

Let us now express the right-hand side of this equation in terms of the mean values characterizing the dielectric medium. From (29.1), the charge on a conductor is connected with the normal component of the displacement vector on its surface in the following way:

$$e = \int \sigma \, dS = \frac{1}{4\pi} \int \mathbf{D} \, d\mathbf{S} \tag{30.6}$$

Substituting this into (30.5) and taking advantage of the constancy of φ on the conductor surface, we obtain

$$dA = \frac{\varphi}{4\pi} \int d\mathbf{D} \, d\mathbf{S} = \frac{1}{4\pi} \int \varphi \, d\mathbf{D} \, d\mathbf{S} \tag{30.7}$$

Now transform the obtained integral over the surface into a space integral:

$$dA = -\frac{1}{4\pi} \int \operatorname{div} (\varphi \, d\mathbf{D}) \, dV \tag{30.8}$$

The minus sign takes into account that in the preceding equation the surface element $d\mathbf{S}$ was directed along the external normal to the conductor, that is, along the internal normal to the surface of the dielectric adjoining the metal.

Making use of (30.2) and (30.3), we expand the divergence under the integral sign in (30.8) as follows:

$$\begin{aligned} \operatorname{div} (\varphi \, d\mathbf{D}) &= (\operatorname{grad} \varphi \cdot d\mathbf{D}) + \varphi \, (\operatorname{div} d\mathbf{D}) \\ &= (\operatorname{grad} \varphi \cdot d\mathbf{D}) + \varphi \, d \, (\operatorname{div} \mathbf{D}) \\ &= -(\mathbf{E} \cdot d\mathbf{D}) \end{aligned} \tag{30.9}$$

From the definition of free energy[2] (8.38),

$$dF = -S\,d\theta + dA = -S\,d\theta + \frac{1}{4\pi}\int (\mathbf{E}\cdot d\mathbf{D})\,dV \qquad (30.10)$$

we find the expression for the increment to its density f in the electric field $\mathbf{E}$:

$$df = \frac{1}{4\pi}(\mathbf{E}\cdot d\mathbf{D}) \qquad (30.11)$$

If the expression for the increment to the free energy density f is known in terms of the electric displacement $\mathbf{D}$, the field $\mathbf{E}$ is determined as follows:

$$\mathbf{E} = 4\pi\left(\frac{\partial f}{\partial \mathbf{D}}\right)_\theta \qquad (30.12)$$

Isotropic Dielectrics. In an isotropic dielectric (gaseous, liquid, or vitreous), when the displacement $\mathbf{D}$ is not too large, the scalar quantity f is linearly expressed in terms of the scalar $\mathbf{D}^2$:

$$f = \frac{\mathbf{D}^2}{8\pi\varepsilon} \qquad (30.13)$$

The dimensionless quantity ε is called the dielectric constant or relative electric permittivity. It depends on the temperature and density of the medium (or its specific volume). Substitution of f into the expression for the field (30.12) yields

$$\mathbf{E} = \frac{\mathbf{D}}{\varepsilon}, \quad \mathbf{D} = \varepsilon\mathbf{E} \qquad (30.14)$$

From this we find the relationship between the field $\mathbf{E}$ and the polarization $\mathbf{P}$:

$$\mathbf{D} = \varepsilon\mathbf{E} = \mathbf{E} + 4\pi\mathbf{P}$$

or

$$\mathbf{P} = \frac{\varepsilon - 1}{4\pi}\mathbf{E} \qquad (30.15)$$

Under the action of an electric field the positive charges move in the direction of the field and the negative away from it. Consequently, the medium is polarized in the direction of the field. Thus Eq. (30.15) shows that in a static field $\varepsilon > 1$. From Eq. (30.13) we can determine the increment to the entropy S' of an isotropic dielectric at $\mathbf{E} \neq 0$. Namely, if the total volume of the dielectric is V, from (8.39) we have

$$S' = -\frac{\partial (fV)}{\partial \theta} = \frac{V\mathbf{D}^2}{8\pi\varepsilon^2}\frac{\partial \varepsilon}{\partial \theta} = \frac{V\mathbf{E}^2}{8\pi}\frac{\partial \varepsilon}{\partial \theta} \qquad (30.16)$$

---

[2] Here temperature is more conveniently expressed in ergs, that is, using the symbol θ.

Differentiation in the latter formula is carried out at a constant electric displacement **D**. Since the displacement vector is determined by the charge on the conductor, the latter is also considered constant in the differentiation, while its potential φ may vary.

For example, let a conductor be at first grounded and a charge e induced on it by bringing up close another charged body. Next the conductor is disconnected from the ground and the inducing charge removed to a sufficient distance. As a result heat is evolved in the dielectric:

$$Q = \theta S'$$

Here the process is reversible in accordance with (8.20) (the *electrocaloric effect*).

If a conductor is connected to another conductor with a sufficiently large capacitance, its potential is conserved. The charge of such a conductor can be changed at constant potential, that is, with the field in the dielectric remaining constant. In these conditions the field and not the electric displacement must be treated as the independent variable. According to the general thermodynamic rules, to go over from the variable **D** to **E** it is necessary to subtract $\mathbf{DE}/4\pi$ from f (see, for example, the transition from energy to enthalpy in Sec. 8). Thus

$$f' \equiv f - \frac{\mathbf{DE}}{4\pi} = -\varepsilon \frac{\mathbf{E}^2}{8\pi}$$

The electrocaloric effect for constant field thus determined has the same value. But this is true only in the case of linear dependence between the field and the electric displacement.

Point Symmetry of Crystals. We shall now examine the dielectric properties of crystals. For this we must introduce certain concepts characterizing crystal symmetry. First of all, solid crystalline bodies are symmetrical with respect to translations in three noncoplanar planes through any values divisible by the lattice constants. This type of symmetry, however, does not concern us at present. More important is that most crystals are symmetrical with respect to rotations through angles $2\pi/n$, where n is an integer equal to 2, 3, 4, 6. Here the crystal facing does not matter: a rotation through the angle n does not affect the crystal's bulk properties, for instance, its dielectric constant. If a crystal is symmetrical with respect to a rotation through $2\pi/n$, it is said to have an *n-fold axis of symmetry*, denoted C_n. Besides axes of symmetry crystals may have planes of symmetry, reflection in which does not affect their bulk properties. Any series of symmetry operations around axes and planes leave at least one point of the crystal in its initial position (the origin of the coordinate system). Therefore such types

of symmetry, as distinct from translational symmetry, are called *point symmetry*.

It will be readily understood why the number n is restricted to the values 2, 3, 4, and 6 if it is taken into account that crystals are characterized by two types of symmetry: point symmetry (with respect to rotations) and translational symmetry. With regard to the latter, a crystal may be likened to parquet flooring. If there is a symmetry axis C_n, the parquet must be laid out in polygons of corresponding symmetry. They may be parallelograms, which repeat themselves in a rotation through an angle of 2π ($n = 1$, there is simply no point symmetry); rectangles symmetrical with respect to a rotation through $2\pi/2$; equilateral triangles with an axis C_3 (rotations through $2\pi/3$); squares and regular hexagons with axes C_4 and C_6 respectively. The discrete character of the translation operation makes for these, and only these, permissible rotations of a crystal in space.

The absence of C_5, C_7 and similar axes of symmetry in natural crystals is indirect proof, based on purely macroscopic properties, of their atomic structure.

Without going into the classification of crystals according to point symmetry, let us consider several examples showing the linear relationships between the field and the electric displacement that occur in them. Here only point symmetry is relevant.

Suppose a crystal has only one preferred axis of symmetry C_n with any possible n, and has no other similar axes at an angle to it; it also has no symmetry plane or axis C_2 perpendicular to the given axis C_n. For example, a regular pyramid with an isosceles triangle as base has a C_3 axis. The plane of symmetry through C_3 is, in the present case, of no consequence.

Unlike an isotropic medium, a medium with such a preferred axis of symmetry can be characterized not only by scalar quantities but also by a vector quantity, provided the vector is directed along the axis. We note that this vector defines precisely the medium, that is, the crystal, and not the external action on it (for example, a force applied to it). Obviously, an essentially isotropic body has no preferred direction.

Free Energy of a Crystal. In this case we have in mind a vector of spontaneous electric polarization. The opposite charges belonging to the crystal are somewhat stretched apart along the preferred axis. Obviously, they cannot on their own accord undergo a displacement in some other direction, since that would violate the given symmetry. Another axis of the same symmetry would be impossible since a displacement of the charges along one axis would destroy the other. A symmetry plane perpendicular to the given axis is also precluded since the charges on either side would be of different

sign. A perpendicular two-fold axis is also impossible because in rotation a plus can not be superimposed on a minus.

But if there is only the axis C_n, a spontaneous polarization vector $\mathbf{P}_0$ exists. Let us now find $f' = f - (\mathbf{E}\cdot\mathbf{D})/(4\pi)$ in such a crystal if the dependence between the field and displacement is linear.

The only scalar that can be linearly constructed from the given vector $\mathbf{P}_0$ and electric field vector $\mathbf{E}$ is the scalar product $(\mathbf{E}\cdot\mathbf{P}_0)$. Furthermore, a quadratic function of the field components of the form $\varepsilon_{ik}E_iE_k$ must be involved in f' instead of $\mathbf{E}^2$. Since f' is a scalar, the quantities ε_{ik} form a tensor of rank 2. Thus

$$f' = -(\mathbf{E}\cdot\mathbf{P}_0) - \frac{1}{8\pi}\varepsilon_{ik}E_iE_k \tag{30.17}$$

If a crystal has no axes of symmetry, spontaneous polarization $\mathbf{P}_0$ may also exist in it in some direction not associated with the crystallographic axes. Therefore (30.17) gives the general expression for f' in a crystalline medium.

Let us now determine the displacement vector. From a formula analogous to (30.12) we obtain

$$D_i = -4\pi\frac{\partial f'}{\partial E_i} = 4\pi P_{0i} + \varepsilon_{ik}E_k \tag{30.18}$$

In the absence of a field a crystal possesses displacement vector $4\pi\,\mathbf{P}_0$.

Such crystals are called *pyroelectric*. The term owes its origin to the following. As a whole, a polarized body possesses a total dipole moment. Consequently, it creates a certain external field. Owing to the attraction of ions from the atmosphere, the field eventually cancels out: negative ions settle on the positive pole of the dipole, and positive ions on the negative pole. But if the body is placed in a flame, the heat changes the spontaneous moment, the field is not cancelled out for a while, and the moment manifests itself. That was how the property was first discovered, hence the name for such crystals (*pyro* is Greek for fire).

Sometimes phase transitions of second order (see Sec. 11) are observed in pyroelectric crystals, in which they turn into nonpyroelectric crystals without any substantial restructuring of the lattice. For instance, if a crystal has a C_n symmetry axis in the pyroelectric phase, a change in temperature or pressure may affect the configuration of its atoms so that a plane of symmetry perpendicular to the symmetry axis appears. Although the change in the lattice takes place smoothly, without a discontinuity, the appearance of a symmetry plane immediately changes the properties of the crystal. Spontaneous polarization becomes impossible. As stated in Sec. 11, at such a transition point it is not the entropy that undergoes a discontinuity but its derivative, specific heat.

The nonpyroelectric phase has no intrinsic dipole moment, near the transition point, but it is easily polarized by an external field because the asymmetric configuration of the atoms corresponding to the polarized state differs but slightly from the initial symmetrical configuration in the absence of a field. Crystals in such a highly polarized state are called *ferroelectrics*. A typical example of a ferroelectric is barium titanate.

The coefficients of the quadratic form in (30.17), ε_{ik}, form the permittivity tensor, which is a symmetric tensor,

$$\varepsilon_{ik} = \varepsilon_{ki} \tag{30.19}$$

according to the definition of a quadratic form.

In a crystal with no elements of point symmetry (axes or planes) ε_{ik}, like any symmetric tensor, has six components. But the appearance of even one two-fold symmetry axis, C_2, or a symmetry plane reduces the number of components of ε_{ik}. Let us direct one of the coordinate axes, say z, along the axis C_2. In a rotation through 180° about the z axis, the x and y coordinates reverse their sign: $x \to -x$, $y \to -y$. The tensor components transform as the products of the respective coordinates:

$$\varepsilon_{xx} \to \varepsilon_{xx}, \qquad \varepsilon_{yy} \to \varepsilon_{yy}, \qquad \varepsilon_{zz} \to \varepsilon_{zz}$$

$$\varepsilon_{xy} \to \varepsilon_{xy}, \qquad \varepsilon_{xz} \to -\varepsilon_{xz}, \qquad \varepsilon_{yz} \to -\varepsilon_{yz}$$

But if the crystal is rotated through a corresponding angle with respect to the symmetry axis, none of its properties can change. In particular, the components of the permittivity tensor must revert to their initial values. Hence ε_{xz} and ε_{yz} are equal to themselves with the opposite sign, or are simply equal to zero.

In the case of one symmetry plane, we make it the x,y-plane. A reflection in this plane changes z to $-z$. Therefore, we again have $\varepsilon_{xz} = -\varepsilon_{xz} = 0$ and $\varepsilon_{yz} = -\varepsilon_{yz} = 0$.

Consider an example of a crystal with one four-fold axis C_4. Obviuosly, a crystal that allows rotations through 90° also possesses symmetry with respect to rotations through $180° = 2 \times 90°$ around the same axis. Thus, $\varepsilon_{xz} = \varepsilon_{yz} = 0$. A rotation through 90° yields $x \to y$, $y \to -x$. Therefore $\varepsilon_{xx} = \varepsilon_{yy}$, $\varepsilon_{xy} = -\varepsilon_{yx} = -\varepsilon_{xy} = 0$. But if the diagonal tensor components in the x,y-plane are the same while the off-diagonal components become zero, the tensor in that plane degenerates into a scalar. When **E** lies in that plane, vector **D** also lies in it and is parallel to **E**.

Let us also consider the case of a crystal with two perpendicular C_2 axes. It will be readily observed that there must then also be a third perpendicular C_2 axis. Indeed, let the first axis undergo the transformation $x \to -x$, $y \to -y$, and the second the transformation $y \to -y$, $z \to -z$. Consecutive application reverts y to its

initial value, and only two coordinates transform: $x \to -x$, $z \to -z$. Comparing this with the case of a single two-fold axis, we find that the C_2 axis along y is responsible for the transformation. Hence, all the off-diagonal components are equal to zero: $\varepsilon_{xy} = \varepsilon_{yz} = \varepsilon_{xz} = 0$. The symmetry axes are the principal axes of the permittivity tensor. In a crystal with no point symmetry the direction of the principal axes of the tensor ε_{ik} is not defined in advance.

If a crystal has two perpendicular four-fold axes of symmetry, they (like two C_2 axes) give rise to a third C_4 axis. In this case a crystal is said to have cubic symmetry. All three diagonal components are the same ($\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{zz}$), while the off-diagonal components become zero. The tensor ε_{ik} degenerates completely into a scalar. The correspondence between the field and the displacement vector is the same as in an isotropic medium. Note that tensor quantities of rank higher than second are not necessarily the same in cubic symmetry crystals as in an isotropic body. A crystal cannot be isotropic in every respect.

A crystal with a three-fold symmetry axis will be examined in Exercise 4.

EXERCISES

1. Two semi-infinite isotropic dielectrics of permittivity ε_1 and ε_2 are separated by a plane; in other words, each fills a half-space on one side of the plane. Within one of the dielectrics there is a point charge e at a distance a from the boundary surface. Determine the field generated by this extraneous charge.

Solution. We employ the method of images. The potential in the medium where the charge is located is determined as in (29.21):

$$\varphi_1 = \frac{e}{\varepsilon_1 r} + \frac{\alpha e}{r'}$$

The potential of the medium with no charge is

$$\varphi_2 = \frac{\beta e}{r}$$

The fields in the first and second media are expressed by the formulas

$$\mathbf{E}_1 = -\operatorname{grad} \varphi_1 = \frac{e\mathbf{r}}{\varepsilon_1 r^3} + \frac{\alpha e\mathbf{r}'}{r'^3}, \qquad E_2 = -\operatorname{grad} \varphi_2 = \frac{\beta e\mathbf{r}}{r^3}$$

The displacement vectors are, accordingly, $\varepsilon_1 \mathbf{E}_1$ and $\varepsilon_2 \mathbf{E}_2$.

The continuity condition for the tangential field components on the boundary surface, where $r = r'$, is given by the equation

$$\frac{1}{\varepsilon_1} + \alpha = \beta$$

The equality of the normal displacement components leads to one more equation:

$$1 - \alpha \varepsilon_1 = \beta \varepsilon_2$$

whence

$$\alpha = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1 (\varepsilon_1 + \varepsilon_2)}, \qquad \beta = \frac{2}{\varepsilon_1 + \varepsilon_2}$$

If the dielectric constant $\varepsilon_2 < \varepsilon_1$ and the charge is in a medium with the constant ε_1, the field of the electric image in the first medium is of the same sign as the field of a charge having the sign of the true charge. Hence, the charge repels from the boundary. For example, in water solutions of electrolytes ($\varepsilon = 81$) ions repel from the surface, and near it the concentration of ions is low (*negative adsorption*).

2. A homogeneous dielectric sphere of radius R is placed in a uniform electric field $\mathbf{E}_0$. Determine the resultant field.

Solution. We assume that within the sphere there is a uniform field $\mathbf{E}_1$ and outside the sphere the field receives the same increment as it would from a dipole moment $\mathbf{d}$ at the centre of the sphere. The condition for the equality of the tangential components of the field and the normal displacement component yields

$$-E_0 + \frac{d}{R^3} = -E_1, \qquad E_0 + \frac{2d}{R^3} = \varepsilon E_1$$

(see Exercise 1, Section 15, and Exercise 6, Section 29). Hence

$$\mathbf{E}_1 = \frac{3\mathbf{E}_0}{\varepsilon + 2}, \qquad \mathbf{d} = \mathbf{E}_0 R^3 \frac{\varepsilon - 1}{\varepsilon + 2}$$

The sphere is polarized uniformly.

Mathematical physics offers proof that any triaxial ellipsoid in an electric field is polarized uniformly. If one of its principal axes is not parallel to the external field, the polarization vector is directed at an angle to it.

3. Calculate the dielectric constant of a gas consisting of dipole molecules with constant moments $\mathbf{d}$.

Solution. Since the potential energy of a dipole in an external field is equal to $-(\mathbf{E} \cdot \mathbf{d})$ (see [16.28]), the portion of the free energy of the gas dependent on the field $\mathbf{E}$ is

$$F_{\mathbf{E}} = -N\theta \ln \int_0^{\pi} \sin \vartheta \, d\vartheta \, e^{(|\mathbf{E}| d \cos \vartheta)/\theta}$$

Here field $|\mathbf{E}|$ can be treated as an external parameter λ (see Sec. 8). It follows from this that $-\partial F_{\mathbf{E}} / \partial |\mathbf{E}| = \Lambda$, where Λ is the mean value of the derivative of the total energy with respect to λ. In this case $\overline{Nd \cos \vartheta}$ (where N is the number of molecules per unit volume). So we can write for the polarization the following formula:

$$P = Nd \, \overline{\cos \vartheta}$$

Consequently,

$$P = N\theta \frac{\partial}{\partial |\mathbf{E}|} \ln \int_0^\pi \sin\vartheta\, d\vartheta\, e^{(|\mathbf{E}|\, d \cos\vartheta)/\theta} = N\theta \frac{\partial}{\partial |\mathbf{E}|} \ln (x^{-1} \sinh x)$$

$$= Nd(\coth x - x^{-1}), \quad x = |\mathbf{E}|\, d/\theta$$

The expression in parentheses is called the Langevin function. In a weak field it takes the form $|\mathbf{E}|d/(3\theta)$. Hence, the polarization of the gas is equal to

$$P = \frac{Nd^2 |\mathbf{E}|}{3\theta}, \qquad \varepsilon = 1 + \frac{4\pi N d^2}{3\theta}$$

From (30.16) the electrocaloric effect of such a gas is negative. When an electric field is switched on, the molecules are oriented mainly along the field. Greater order corresponds to smaller entropy. Consequently, heat must isothermally dissipate into the external medium, as in isothermal compression of a gas.

If the molecules have no intrinsic dipole moments, in the first approximation the polarization does not depend on temperature. Its value is much less than in gases with dipole moments.

4. Show that in a crystal with a three-fold symmetry axis the permittivity tensor degenerates into a scalar in a plane perpendicular to the axis.

Solution. Introduce the complex coordinates $\xi = x + iy$, $\eta = x - iy$. In the most general case the tensor ε_{xx}, $\varepsilon_{xy} = \varepsilon_{yx}$, ε_{yy} receives three complex components: $\varepsilon_{\xi\xi}$, $\varepsilon_{\xi\eta} = \varepsilon_{\eta\xi}$, $\varepsilon_{\eta\eta}$. In a rotation through 120° the coordinate ξ receives the factor $e^{2\pi i/3}$, and $\varepsilon_{\xi\xi}$ receives the factor $e^{4\pi i/3}$. Therefore $\varepsilon_{\xi\xi} = e^{4\pi i/3}\, \varepsilon_{\xi\xi}$, whence $\varepsilon_{\xi\xi} = 0$. Similarly, $\varepsilon_{\eta\eta} = e^{-4\pi i/3}\, \varepsilon_{\eta\eta}$ and $\varepsilon_{\eta\eta} = 0$. There remains the component $\varepsilon_{\xi\eta} = \varepsilon_{\eta\xi}$, which is multiplied by 1. But $\varepsilon_{\xi\eta}$ is a real quantity, since $\overset{*}{\varepsilon}_{\xi\eta} = \varepsilon_{\xi^*\eta^*} = \varepsilon_{\eta\xi} = \varepsilon_{\xi\eta}$. The permittivity tensor in a plane perpendicular to the axis is given by one real number, that is, it degenerates into a scalar.

31

DIRECT CURRENT

Basic Equations. Direct current in a conductor can be visualized as continuously increasing polarization. Indeed, the dipole moment of a unit volume of a neutral medium is the following sum over all the charges:

$$\mathbf{P} = \sum_{+,\,-} \rho\,(\mathbf{r}_+ - \mathbf{r}_-) \tag{31.1}$$

Hence

$$\frac{\partial \mathbf{P}}{\partial t} = \sum_{+,\,-} \rho\,(\mathbf{v}_+ - \mathbf{v}_-) \qquad (31.2)$$

But this is the current passing in unit time across unit surface, or the *current density* $\mathbf{j}$. Thus, for a constant electric field Eq. (28.28), taking into account (28.23) and (31.2), assumes the form

$$\operatorname{curl} \mathbf{H} = \frac{4\pi}{c}\frac{\partial \mathbf{P}}{\partial t} = \frac{4\pi \mathbf{j}}{c} \qquad (31.3)$$

At constant magnetic induction, (28.27) also yields

$$\operatorname{curl} \mathbf{E} = 0 \qquad (31.4)$$

It follows from Eq. (31.3) that

$$\operatorname{div} \mathbf{j} = \frac{c}{4\pi} \operatorname{div} \operatorname{curl} \mathbf{H} = 0 \qquad (31.5)$$

This means that the lines of the current density are closed (like magnetic induction lines). At the boundary between two media the normal components of the current density vector are conserved:

$$j_{n1} = j_{n2}$$

To satisfy Eq. (31.4) the electric field should be represented via a potential:

$$\mathbf{E} = -\operatorname{grad} \varphi \qquad (31.6)$$

Unlike the electrostatic potential in a conductor through which current is flowing, the potential φ represents a variable quantity dependent on the coordinates.

Ohm's Law. All conductors are classified into two types. In conductors of the first type the charge is transferred by electrons and no movement of matter occurs. To this type belong all metals and semiconductors. In conductors of the second type the charge is transferred by ions, that is, movement of the matter itself occurs. An example are solutions of electrolytes.

In a sufficiently weak field in a homogeneous and isotropic medium the proportional dependence

$$\mathbf{j} = \sigma \mathbf{E} \qquad (31.7)$$

always holds. It is called *Ohm's law*.

The factor σ is called the *electrical conductivity* (specific conductance). It has the dimension s^{-1}. (In this section the notation σ refers only to conductivity, as distinct from Section 29, where σ denotes the surface-charge density.) For a single crystal of a metal

or semiconductor the dependence (31.6) is of tensor form:

$$j_i = \sigma_{ik} E_k \tag{31.8}$$

We shall not use this more general form of Ohm's law, however, assuming solids to be polycrystals, as they in fact usually are.

The main difference between metals and semiconductors is that the former's conductivity increases as temperature decreases, whereas the latter's conductivity decreases and at absolute zero tends to zero. Also the conductivity of semiconductors is many orders of magnitude less than that of "good" metals, such as copper, silver, or gold. The value of σ for metals at room temperature is of the order of 10^{-18} s^{-1}. Examples of semiconductors are germanium, silicon, and cuprous oxide. The conductivity of semiconductors usually has no definite value and depends greatly on manufacturing techniques and admixtures (see Sec. 43).

For metals Ohm's law holds in all experimentally attainable fields. Semiconductors display appreciable deviations from field-current proportionality. The field increases the number of conduction electrons and changes the conditions of their passage through the crystal lattice.

If Ohm's law is applicable to a conductor, or there is an established dependence between the field and the current density, the set of equations describing the flow of direct current is closed. When $\mathbf{j} = \sigma \mathbf{E}$,

$$\operatorname{div} \mathbf{j} = -\operatorname{div} (\sigma \operatorname{grad} \varphi) \tag{31.9}$$

which for constant conductivity reduces to the Laplace equation

$$\nabla^2 \varphi = 0 \tag{31.10}$$

On the boundary between two conductors the continuity condition of the normal component of the current, which is written in terms of potential gradients as

$$\sigma_1 \operatorname{grad}_{n1} \varphi = \sigma_2 \operatorname{grad}_{n2} \varphi \tag{31.11}$$

holds. On the boundary between a conductor and a dielectric the normal component of grad φ vanishes since $j_n = 0$.

Joule Heat. From [14.32], the work done by an electric field on moving charges in unit time is

$$\frac{dA}{dt} = \sum \mathbf{E}\rho (\mathbf{v}_+ - \mathbf{v}_-) \tag{31.12}$$

Going over to current density with the help of (31.2), we obtain

$$\frac{dA}{dt} = \mathbf{Ej} \tag{31.13}$$

If the physical state of the conductor does not change, the energy received by its particles must be dissipated in the form of heat,

that is, $dA/dt = dQ/dt$. When Ohm's law (31.7) is applicable to a conductor, we can write the following relationship:

$$\frac{dQ}{dt} = \sigma \mathbf{E}^2 \tag{31.14}$$

The sign of the evolved heat (which is called the *Joule heat*) does not change when the direction of the field changes. Therefore the evolution of heat in the ohmic resistance of a conductor is an irreversible process, like viscous friction (Sec. 17). It is accompanied by an increase in the entropy of the system comprising the conductor and the surrounding medium.

In conductors of the first type the rate of increase in entropy is connected with the Joule heat by the same relationship as in reversible heat transfer:

$$\frac{dS}{dt} = \int \frac{Q}{\theta} dV = \int \frac{\sigma \mathbf{E}^2}{\theta} dV \tag{31.15}$$

Unlike the general relationship (8.20), this formula has the equality sign, because in a metal or semiconductor no irreversible changes in the parameters of the system occur. Given constant physical conditions, the state of a conductor of the first type through which current is passing remains absolutely constant. Since entropy increases, the conductivity σ is always positive.

In conductors of the second type, in which matter is transferred, thereby changing the concentration of the components, Eq. (31.15) cannot hold. Irreversibility in the case of variable concentration is due not only to the evolution of heat but also to diffusion processes (see Sec. 17).

Total Current in a Conductor. We shall denote the total current passing across a cross section of a conductor by I:

$$I = \int \mathbf{j} \, d\mathbf{S} \tag{31.16}$$

Consider two cross sections (Figure 43) separated by side surfaces through which no current passes (the current lines lie on them). By virtues of the charge conservation law the integral (31.16) is the same for both sections.

The Joule heat evolved in the volume between the sections is

$$\frac{dQ}{dt} = \int \mathbf{jE} \, dV = -\int (\mathbf{j} \cdot \operatorname{grad} \varphi) \, dV \tag{31.17}$$

Transforming this integral by parts, we obtain

$$\frac{dQ}{dt} = -\int \varphi \mathbf{j} \, d\mathbf{S} + \int \varphi \, (\operatorname{div} \mathbf{j}) \, dV \tag{31.18}$$

Since div $\mathbf{j} = 0$, the volume integral is equal to zero. Now choose sections *1* and *2* on the equipotential surfaces. Taking into account

that in Eq. (31.18) $d\mathbf{S}$ is everywhere directed along the external normal, we obtain the following expression for the Joule heat:

$$\frac{dQ}{dt} = (\varphi_1 - \varphi_2) \int \mathbf{j}\, d\mathbf{S} = (\varphi_1 - \varphi_2)\, I \tag{31.19}$$

This quantity is positive, since the current flows from higher potential to lower. By virtue of the linearity of the equations it is proportional to the potential difference:

$$I = \frac{\varphi_1 - \varphi_2}{R} \tag{31.20}$$

The coefficient R is called the *resistance* of the conductor between the equipotential sections *1* and *2*.

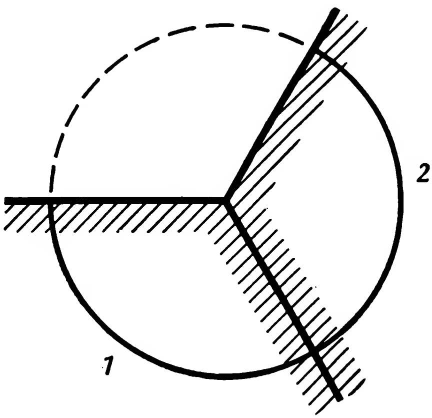


Figure 43

If the conductor is linear or, more precisely, cylindrical, of length l and cross-sectional area F, then $I = Fj$, $\varphi_1 - \varphi_2 = |\,\mathbf{E}\,|\, l$. From this, with the help of the expression (31.16), resistance is expressed in terms of conductivity as follows: $R = l/(\sigma F)$. Finally, from (31.19) and (31.20) we obtain

$$\frac{dQ}{dt} = \frac{(\varphi_1 - \varphi_2)^2}{R} = RI^2 \tag{31.21}$$

The Galvanic Cell. It was shown in Section 29 that in a broken circuit consisting of several metals the potential on the ends is the same if the ends are of the same metal. This is because the nature of the current carrier in all metals is the same. But if a circuit includes, in addition to metals, an electrolyte, that is, a conductor through which charge is carried by ions, a potential difference appears at the ends of an open circuit.

Consider the following circuit: zinc, an electrolyte consisting of solutions of zinc sulfate $ZnSO_4$ and copper sulfate $CuSO_4$, and copper. When an atom of zinc passes into the solution, it becomes a cation Zn^{++}, and a charge passes to the zinc electrode in the form of two electrons. Further, one ion of copper precipitates out of the electrolyte onto the copper electrode, which thus acquires two excess electric charges (it is short of two electrons). In a broken circuit one end of which has a positive charge and the other a negative one, there is a potential difference. It is equal to the *electromotive force* (emf) of the galvanic cell, $\mathscr{E}$.

If we now join the electrodes with a conductor, a current appears in the circuit. It is maintained by virtue of the fact that more free energy evolves per atom in the dissolution of the zinc than is required for the precipitation of the copper. The process is reversible: by passing current in the reverse direction the copper can be made to dissolve and the zinc to precipitate.

Since the reaction in a galvanic cell takes place at constant temperature and pressure we must take not the free energy but the thermodynamic potential G (see Sec. 8). Its change per particle is the chemical potential of the substance entering the reaction. It follows from the reversibility of the process that the work done on the charges in the passage of current is equal to the total change in the thermodynamic potential of the system.

Dissolution of zinc yields the work per atom, $\mu_{Zn} - \mu_{ZnSO_4}$, and the precipitation of the copper, the work $\mu_{CuSO_4} - \mu_{Cu}$. The resulting change in thermodynamic potential is equal to

$$\delta G = \mu_{Zn} - \mu_{ZnSO_4} - \mu_{Cu} + \mu_{CuSO_4} \tag{31.22}$$

In the process, two elementary charges pass through the short-circuited galvanic cell. By definition, work is then done that is equal to the potential difference multiplied by the transferred charge. Hence the emf of a galvanic cell is

$$\mathscr{E} = \frac{\delta G}{2|e|} \tag{31.23}$$

Let us now find the relationship between the emf and the current passing through the cell. The potential difference at the ends of the ith conductor in the circuit is $\varphi_{1i} - \varphi_{2i}$. It is related to the current by the relationship (31.20):

$$\varphi_{1i} - \varphi_{2i} = R_i I$$

Adding up such equations for all the conductors in the circuit, we obtain the emf on the left and $I \sum R_i$ on the right. Hence

$$I = \frac{\mathscr{E}}{\sum R_i} \tag{31.24}$$

In the process the heat of the chemical reaction, Q, evolves in the cell. It is connected with the electromotive force by a formula similar to (13.9).

Thermoelectromotive Force. The emf concept is applicable not only to galvanic cells. In the most general case the emf is equal to the work done on a unit charge passing through a closed circuit:

$$\mathscr{E} = \int \mathbf{E}\, d\mathbf{l} \tag{31.25}$$

Let us examine the emf generated in a system of conductors by a temperature gradient. But first several definitions.

The passage of current represents an irreversible relaxation process of approach to statistical equilibrium. If a potential difference has been created in a conductor, nonequilibrium conditions appear in it, with the electric current performing the approximation to equilibrium. Equilibrium should not be confused with steady-state nonequilibrium conditions, for example the flow of direct current from an external emf source.

Similarly, a temperature gradient disturbs thermal equilibrium and is levelled out by heat flow: this is another relaxation process of approach to statistical equilibrium.

These relaxation processes are interrelated: a temperature gradient produces electric current in a conductor, while a potential gradient produces a heat flux. There exist certain relationships between these two *coupled processes*, which shall be established here.

For simplicity we shall consider only linear conductors. Let the coordinate of a point laid off along a conductor be x. Then the expression for the current, given temperature and potential gradients, is written as follows:

$$I = \frac{1}{R_0}\left(E_i - \alpha\,\frac{\partial\theta}{\partial x}\right) = -\frac{1}{R_0}\left(\frac{d\varphi}{dx} + \alpha\,\frac{d\theta}{dx}\right) \tag{31.26}$$

where R_0 is the resistance per unit length of the conductor. This equation should be seen as a definition of the *coupling coefficient* α.

Now consider a circuit consisting of two metals, I and II (Figure 44). The metals are soldered at two points (junctions), which are at different temperatures θ_1 and θ_2. Let us show that emf is generated in the circuit and is expressed in terms of the α' s of both metals.

The emf is equal to the potential difference at the ends of the open circuit, which coincides with Eq. (31.25):

$$\mathscr{E} = \int E_x\, dx = -\int \frac{d\varphi}{dx}\, dx = \varphi_1 - \varphi_2$$

We assume that the end points are separated by a break in the circuit, as shown in Figure 44. At $I = 0$ from (31.26) we obtain

$$\mathscr{E} = \int E_x\,dx = \int \alpha \frac{d\theta}{dx}\,dx \tag{31.27}$$

This integral is taken from one end of the broken circuit to the other. It is convenient to go over to the integration variable θ.

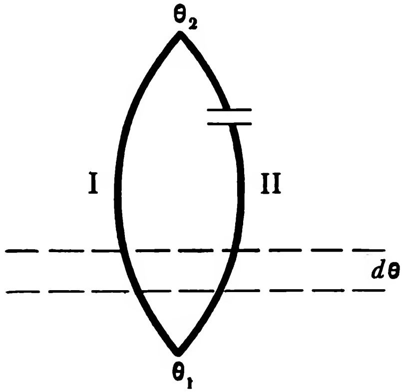


Figure 44

Since in passing through the whole of the closed circuit any temperature gradient $d\theta$ occurs in opposite directions in conductors I and II,

$$\mathscr{E} = \int_{\theta_1}^{\theta_2} (\alpha_{\mathrm{I}} - \alpha_{\mathrm{II}})\,d\theta \tag{31.28}$$

From this equation it can be seen why two metals are needed to obtain a thermo-emf in a circuit: in a ring consisting of one metal $\alpha_{\mathrm{I}} - \alpha_{\mathrm{II}} = 0$. Furthermore, there must be a temperature difference at the junctions ($\theta_1 \neq \theta_2$). This is in agreement with the second law of thermodynamics (Sec. 8): to obtain work from a heat engine a temperature difference is needed.

If the ends of the broken circuit are at potentials φ_1 and φ_2, the emf is $\varphi_1 - \varphi_2$. For an infinitely small temperature difference at the junctions we find the expression for the differential of the thermo-emf:

$$d\varphi = (\alpha_{\mathrm{I}} - \alpha_{\mathrm{II}})\,d\theta \tag{31.29}$$

Peltier Effect. We shall now consider another "coupled" phenomenon, the transfer of heat in a conductor in a nonzero electric field. First we shall find the expression for the energy flux in a conductor

in the presence of potential and temperature gradients. A charge e at a point with a potential φ possesses an energy $e\varphi$. Therefore in unit time a current I transfers an amount of energy $I\varphi$. If there is a temperature gradient $d\theta/dx$, there is a heat flux $-\gamma d\theta/dx$. The minus indicates that heat is transferred in the direction of decreasing temperature. Finally, there is an energy flux of a "coupled" origin, which we shall denote βE_x. Denoting the total energy flux along the conductor by W, we obtain

$$W = I\varphi + \beta E_x - \gamma \frac{d\theta}{dx} \tag{31.30}$$

We further express the electric field in terms of the current according to Eq. (31.26). Then

$$W - \varphi I = \beta R_0 I + (\alpha\beta - \gamma)\frac{d\theta}{dx} \tag{31.31}$$

Suppose now that there is no temperature gradient, and apply Eq. (31.31) to the junction of the two metals. The current at the junction, I, is continuous, while the potential suffers a discontinuity equal to the contact difference $\varphi_2 - \varphi_1$. Developing the difference between the energy flowing to the junction in unit time and energy flowing away from it, we have

$$(W - \varphi_2 I) - (W - \varphi_1 I) = (\beta_2 R_{02} - \beta_1 R_{01})\, I \tag{31.32}$$

where W is the total energy flux in the conductor, and $I\varphi$ is the energy flux of the charges in the electric field. This part of the energy flux can, in principle, be turned into mechanical work. Therefore the difference $W - \varphi I$, according to the first law of thermodynamics (8.9), has the meaning of heat flow. If the quantity $W - \varphi I$ undergoes a discontinuity at some point, evolution or adsorption of heat occurs there. This is known as the *Peltier effect*.

The difference $\beta_2 R_{02} - \beta_1 R_{01}$ is conventionally called the *Peltier coefficient* for the given junction and is denoted by the symbol $\Pi_{\text{II-I}}$. We can also speak of separate coefficients, Π_{II} and Π_{I}:

$$\Pi_{\text{II-I}} = \Pi_{\text{II}} - \Pi_{\text{I}} = \beta_2 R_{02} - \beta_1 R_{01} \tag{31.33}$$

The Peltier effect is a linear function of the electric current. Therefore, unlike Joule heat, *Peltier heat* evolves reversibly: when the direction of the current changes, its sign reverses. This assertion, however, requires a stricter proof. It was mentioned before that states with a nonzero current are characterized by steady-state, and not equilibrium, conditions. Nevertheless, we shall assume the Peltier effect to be reversible. Having made this assumption, we apply the second law of thermodynamics to the system of conductors in Figure 44. Let us consider it as a reversible heat engine that operates at a temperature difference $d\theta$ between the heat source and

heat sink. Somewhat more heat is conducted to the hot end than to the cold. According to the definition of the Peltier coefficient, the heat supply to the junction is equal to $-\Pi_{II\text{-}I}I$. But since heat is drained from the other junction, in a reversible process only a portion of the heat $\Pi_{II\text{-}I}I$ can be turned into useful work. That portion is equal to the ratio $d\theta/\theta$. That, according to the second law of thermodynamics (8.25), is the efficiency of a reversible heat engine operating at a temperature difference $d\theta$ between the heat source and the heat sink.

Work is expended on maintaining the current produced by the thermo-emf. Therefore from (31.29) we obtain

$$-\Pi_{II\text{-}I}I\frac{d\theta}{\theta}=(\alpha_I-\alpha_{II})\,I\,d\theta \tag{31.34}$$

From this we find the relationship between the coupling coefficients:

$$\frac{\Pi_{II}-\Pi_I}{\theta}\equiv\frac{\beta_2R_{02}-\beta_1R_{01}}{\theta}=\alpha_{II}-\alpha_I \tag{31.35}$$

But since metals I and II were chosen arbitrarily, and so was the temperature of the medium, an equation of the form (31.35) should hold for each metal separately. Discarding the subscripts, we obtain

$$\beta R_0=\alpha\theta \tag{31.36}$$

This formula was obtained in the nineteenth century by W. Thomson, who assumed the reversibility of the Peltier effect. Strict proof was offered in 1934 by L. Onsager for coupling coefficients in linear relationships of the type (31.26) and (31.30) (see Sec. 40).

Onsager showed, in particular, that the electrical conductivity tensor σ_{ik} in crystalline conductors is symmetric: if a unit potential gradient along the x axis generates a current along the y axis, then a unit potential gradient along the y axis generates an identical current along the x axis.

Thomson Effect. Thomson, on the basis of his theory of thermoelectric phenomena, predicted one more effect, which for quite some time could not be detected experimentally: in passing along a non-uniformly heated conductor current evolves additional heat besides the Joule heat.

Let us calculate the derivative of the energy flux with respect to the coordinate after substituting the expression for E_x from Eq. (31.26) into it:

$$\frac{dW}{dx}=I\,\frac{d\varphi}{dx}+I\,\frac{d}{dx}\,\beta R_0+\frac{d}{dx}\,(\alpha\beta-\gamma)\,\frac{d\theta}{dx} \tag{31.37}$$

If we substitute $E_x = -d\varphi/dx$ from (31.26) and the coefficient α from (31.16), dW/dx appears as a sum of three terms:

$$\frac{dW}{dx} = -I^2R_0 + \frac{d}{dx}(\alpha\beta-\gamma)\frac{d\theta}{dx} + I\frac{d\theta}{dx}\left(\frac{d}{d\theta}\beta R_0 - \frac{\beta R_0}{\theta}\right)$$

The change in the energy flux over unit length in steady-state conditions is equal to the energy carried away from unit length of the conductor. The first term expresses the Joule heat. The minus sign indicates that it is dissipated. The second term is associated with the change in purely thermal flow along the conductor. The third term expresses the additional heat due to the combined action of the current and the temperature gradient. Obviously, this additional energy cannot be dissipated in any other form but heat. In somewhat changed form the expression for this quantity of heat is

$$\left(\theta\frac{d}{d\theta}\frac{\Pi}{\theta}\right) I\frac{d\theta}{dx} \tag{31.38}$$

The proportionality factor in parenthesis is called the *Thomson coefficient.*

EXERCISES

1. A medium with low conductivity σ fills a half-space. Immersed in it are two spherical electrodes of radius r_0 each. The conductivity of the electrodes is much greater than that of the medium. Show that the current between the electrodes is the same as that flowing from a separate electrode whose potential with respect to infinitely remote points of the medium is equal to the potential difference between the electrodes.

Solution. Let the potential of a separate electrode be φ_0. It can be treated as constant over the electrode since the conductivity of the medium is assumed much smaller than the conductivity of the electrode material. At a distance r from the centre of the electrode the potential in the medium is φ_0r_0/r. Hence the current passing into the medium can be expressed as follows:

$$I = 2\pi r^2\sigma\varphi_0\frac{r_0}{r^2} = 2\pi\sigma\varphi_0r_0$$

The potential difference of the two electrodes is equal to

$$\varphi = \frac{\varphi_0r_0}{2r_1} - \frac{\varphi_0r_0}{2r_2}$$

where r_1 and r_2 are the distances of the given point from the centres of the electrodes. The current between them is best calculated in terms of the current passing through the median plane separating them. Introducing for the time being the distance between them, equal to $2a$, we find that the

normal component of the current density on the plane is equal to

$$j_n = \frac{\sigma\varphi_0 r_0 a}{(a^2+\rho^2)^{3/2}}$$

where ρ is the distance of the point from the line joining the electrodes.

The total current is

$$I = \int j_n \pi\rho \, d\rho = 2\pi\sigma\varphi_0 r_0$$

Since $\operatorname{div} \mathbf{j} = 0$ the current across any surface separating the electrodes is the same and is independent of a.

This problem explains the principle of grounding: for telephone communication, for example, one wire is sufficient, with the ground acting as the second, since the resistance in it does not depend on the distance between the electrodes inserted in the ground. Ground currents can be amplified to tap conversations, which is why two-wire communication is nevertheless employed.

2. Express the heat of a chemical reaction in a galvanic cell in terms of its electromotive force.

32

MAGNETIC PROPERTIES OF NONFERROMAGNETIC MEDIA

Work Done by a Magnetic Field. Like the electrical properties of dielectrics, the magnetic properties of various media are conveniently described with the help of the expression for the free energy of media in a magnetic field. Suppose a magnetic field is due to a certain distribution of the current density $\mathbf{j}$. As is known, only an electric field does direct work on charges [14.32]. But if the field is due to induction, the work done on electric charges can also be expressed in terms of the change in the magnetic field according to Eq. (28.27).

Using this equation we calculate the change in the energy of a magnetic field in time dt. If the work is done on currents, it is convenient to define it with a minus sign with respect to the energy of the field. Therefore

$$dA = -dt \int (\mathbf{E}\cdot\mathbf{j}) \, dV \tag{32.1}$$

We substitute the current density according to equation (31.3) to get

$$dA = -\frac{c\, dt}{4\pi} \int (\mathbf{E}\cdot \operatorname{curl} \mathbf{H}) \, dV \tag{32.2}$$

Now we transform the obtained integral by parts, taking into account that on an infinitely distant surface the field becomes zero:

$$dA = -\frac{c\,dt}{4\pi}\int \mathbf{E}(d\mathbf{S}\times\mathbf{H}) + \frac{c\,dt}{4\pi}\int \mathbf{E}(\nabla_{\mathbf{E}}\times\mathbf{H})\,dV$$
$$= -\frac{c\,dt}{4\pi}\int (\mathbf{H}\times\nabla_{\mathbf{E}})\,\mathbf{E}\,dV = -\frac{c\,dt}{4\pi}\int (\mathbf{H}\cdot\mathrm{curl}\,\mathbf{E})\,dV \quad (32.3)$$

We finally replace curl **E** according to Eq. (28.27) and cancel out dt to get

$$dA = \frac{1}{4\pi}\int (\mathbf{H}\cdot d\mathbf{B})\,dV \quad (32.4)$$

Reasoning as in Section 30, we see that the differential of the density of the free energy due to the magnetic field is equal to

$$df_{\mathrm{m}} = \frac{1}{4\pi}(\mathbf{H}\cdot d\mathbf{B}) \quad (32.5)$$

Note that this expression is analogous to f', involving the differential of the electric field; that is because magnetic induction is the mean value of the magnetic field, which corresponds to the definition of electric field and not of the displacement vector in a medium.

Thus the differentials of similar electrical and magnetic quantities have the form (introducing the subscript "e")

$$4\pi\,df_{\mathrm{e}} = (\mathbf{E}\cdot d\mathbf{D}), \qquad 4\pi\,df'_{\mathrm{m}} = -(\mathbf{B}\cdot d\mathbf{H})$$
$$4\pi\,df'_{\mathrm{e}} = -(\mathbf{D}\cdot d\mathbf{E}), \qquad 4\pi\,df_{\mathrm{m}} = (\mathbf{H}\cdot d\mathbf{B})$$

Magnetic Permeability. We see that to establish the connection between magnetic induction and magnetic field we must calculate f_{m} as a function of induction and then determine **H** according to the formula

$$\mathbf{H} = 4\pi\frac{\partial f_{\mathrm{m}}}{\partial\mathbf{B}} \quad (32.6)$$

Experience shows that in all media except ferromagnetics, which will be considered in the next section, the magnetic field and magnetic induction are proportional. Therefore the dependence of the free energy density on magnetic induction must be sought in quadratic form

$$f_{\mathrm{m}} = \frac{\mathbf{B}^2}{8\pi\mu} \quad (32.7)$$

Substituting (32.7) into (32.6), we obtain

$$\mathbf{H} = \frac{\mathbf{B}}{\mu}, \quad \text{or} \quad \mathbf{B} = \mu\mathbf{H} \quad (32.8)$$

The coefficient μ is called the *relative magnetic permeability* of the medium. Unlike relative dielectric permittivity ε, magnetic permeability may be either less or greater than unity. In the former case the substance is called *diamagnetic*, in the latter *paramagnetic*.

Paramagnetic substances are those whose molecules possess angular momenta. In a magnetic field the net angular momentum assumes an orientation, corresponding to thermodynamic equilibrium, parallel to the field. Then the polarization vector is in the same direction as **B**, and consequently $B > H$ and $\mu > 1$. But in addition to this effect currents appear in the molecules which, according to Lenz's induction law, weaken the external field. The contribution of these currents to magnetic permeability is in general smaller than the contribution of the angular momenta of the molecules. An evaluation will be carried out later.

If, however, there is no net angular momenta, the polarization vector due to induction currents is directed opposite the external field, and $\mu < 1$. This is what produces diamagnetism.

The Van Leeuwen Theorem. The magnetic properties of bodies are in final analysis of a purely quantum nature. Indeed, the magnetic field is not involved in the classical partition function

$$Z = \int e^{-\mathscr{H}(\mathbf{p},\, \mathbf{r})/\theta}\, d\Gamma$$

In a magnetic field, all the momenta $\mathbf{p}$ of charged particles are, as we know from [14.24], replaced by $\mathbf{p}' = \mathbf{p} - e\mathbf{A}/c$. If the $\mathbf{p}'$ are taken as new independent integration variables, the phase volume element $d\Gamma = \prod_i dp_i\, dx_i$ is replaced by $\prod_i dp'_i\, dx_i$, the Jacobian of the transformation being equal to unity. This is seen from the fact that the vector itself depends only on the coordinates. Consequently, the vector potential is not involved in Z:

$$Z = \int e^{-\mathscr{H}(\mathbf{p},\, \mathbf{r})/\theta} \prod_i dp_i\, dx_i = \int e^{-\mathscr{H}(\mathbf{p}',\, \mathbf{r})/\theta} \prod_i dp'_i\, dx_i$$

All constants describing the magnetic properties of media are in one way or another dependent on Planck's constant.

Free Energy of a Substance in Magnetic Field. In subsequent calculations we shall be assuming the magnetic field to be weak so as to obtain the required general expressions. For dia- and paramagnetic substances this assumption is fully justified. In an external field, the supplementary energy of an atom or molecule is of the order of the product of the Bohr magneton [33.49] multipled by H. Even if $H \sim 10^5$ gauss, the energy will be of the order of 10^{-15} erg.

This is in any case very small compared to the atomic energy scale ($\sim 10^{-12}$ erg). Since at room temperature $\theta \sim 4 \times 10^{-14}$ erg, the magnetic increment is small compared to the energy of thermal motion.

Let us calculate the increment to the energy of the ground state of a quantum system placed in a uniform magnetic field **H**. From [17.26] the vector potential of the field is equal to

$$\mathbf{A} = \frac{1}{2} \mathbf{H} \times \mathbf{r} \tag{32.9}$$

Therefore, in the nonrelativistic approximation [13.38] the Hamiltonian operator is written as

$$\hat{\mathscr{H}} = \sum_i \left[\frac{1}{2m} \left(\hat{\mathbf{p}}_i - \frac{e}{2c} \mathbf{H} \times \mathbf{r}_i \right)^2 + U(r_i) - \frac{eh}{mc} (\mathbf{H} \cdot \hat{\sigma}_i) \right] \tag{32.10}$$

We note that the hat over $\mathscr{H}$ and **p** denotes an operator in the quantum mechanical sense. The symbols $\hat{\sigma}_i$ denote the Pauli operators of the electrons [30.31][3]. Squaring the expression inside the brackets in the first term, we obtain

$$\hat{\mathscr{H}} = \sum_i \left\{ \frac{\hat{\mathbf{p}}_i^2}{2m} + U(r_i) - \frac{e}{4mc} [\hat{\mathbf{p}}_i (\mathbf{H} \times \mathbf{r}_i) + (\mathbf{H} \times \mathbf{r}_i) \hat{\mathbf{p}}_i] + \frac{e^2}{8mc^2} (\mathbf{H} \times \mathbf{r}_i)^2 - \frac{eh}{mc} (\mathbf{H} \cdot \hat{\sigma}_i) \right\} \tag{32.11}$$

Transposing the multiplicands in the mixed vector products, we express them in terms of the mechanical moment operators [Sec. 24]:

$$\hat{\mathbf{p}}_i (\mathbf{H} \times \mathbf{r}_i) + (\mathbf{H} \times \mathbf{r}_i) \hat{\mathbf{p}}_i = 2\mathbf{H} (\mathbf{r}_i \times \hat{\mathbf{p}}_i) = 2\mathbf{H}\hat{\mathbf{M}} \tag{32.12}$$

The noncommutativity of the operators $\hat{\mathbf{p}}$ and **r** does not affect the vector product because its components involve only different projections of **p** and **r**.

Now let the z axis be directed along the magnetic field. The Hamiltonian is then equal to

$$\hat{\mathscr{H}} = \hat{\mathscr{H}}_0 - \frac{ehH}{2mc} (\hat{L}_z + 2\hat{S}_z) + \sum_i \frac{e^2 H^2}{8mc^2} (x_i^2 + y_i^2) \tag{32.13}$$

Here $\hat{\mathscr{H}}_0$ is the Hamiltonian not disturbed by the field, $\hat{L}_z$ is the operator of the projection on the z axis of the total orbital angular momentum expressed in units of h, and $\hat{S}_z$ is the operator of the

---

[3] Note that the ratio of the magnetic moment to the mechanical moment is twice as great for spin as for orbital angular momentum.

projection of the spin angular momentum on the z axis. The third term involves the square of a vector product:

$$(\mathbf{H} \times \mathbf{r})^2 = H^2 r^2 - (\mathbf{H} \cdot \mathbf{r})^2 = H^2 r^2 - H^2 z^2 = H^2 (x^2 + y^2) \tag{32.14}$$

The corrections to the energy, that is, to the eigenvalues of the Hamiltonian $\hat{\mathscr{H}}_0$, must be taken into account in the first and second approximations of the perturbation theory [Sec. 32]. In the first approximation the correction is equal to the mean value of the perturbation Hamiltonian with respect to the unperturbed state; this refers to both terms of the perturbation in (32.13), linear and quadratic (with respect to H).

In principle it would be right to also take account of the second-approximation correction with respect to the Hamiltonian, linear with respect to the magnetic field. But as is known from [Sec. 32], in the denominator such a correction involves the differences of the energy eigenvalues of the unperturbed system. If the system has no total angular momentum, these differences, as was pointed out, are of the order of 10^{-12}, so that the second approximation makes a very small contribution to the energy being determined. If the perturbed state has a total angular momentum and therefore has a fine structure, the energy differences in the denominator may not be great. But in that case the terms in the energy eigenvalues quadratic with respect to H need not be considered at all in calculating *paramagnetic susceptibility*[4], since their contribution is very small.

Consequently, it is always possible to confine oneself to the mean value of the perturbing energy with respect to the unperturbed state:

$$\Delta E = -\frac{ehH}{2mc} (\langle L_z \rangle + 2 \langle S_z \rangle) + \frac{e^2 H^2}{8mc^2} \langle \sum x_i^2 + y_i^2 \rangle \tag{32.15}$$

Here the only term quadratic with respect to the magnetic field is the third term in (32.13). The quantities $\langle L_z \rangle$ and $\langle S_z \rangle$ are the mean values of the projections of the total angular orbital and spin momenta. The angle brackets denote quantum-mechanical means (see [25.19]). The increment to the free energy per one atom due to the magnetic field is (see Sec. 7)

$$\Delta F = -\theta \ln \sum e^{-\Delta E/\theta} = -\theta \ln \sum e^{-\beta H (\langle L_z + S_z \rangle)/\theta}$$

$$+ \theta \ln \left\{ \exp \left[-\frac{e^2 H^2}{8mc^2 \theta} \langle \sum (x_i^2 + y_i^2) \rangle \right] \right\} \tag{32.16}$$

Here $\beta = eh/(2mc)$ is the Bohr magneton. The summation is over all projections of angular momenta.

---

[4] Magnetic susceptibility is the proportionality coefficient between **M** and **H**.

Diamagnetism. We shall start with the case when in the ground state of a system the magnetic moment is zero. Then the first term in parentheses in (32.15) is unity and the free energy receives the increment

$$(\Delta F)_{\text{diamagnetic}} = \frac{e^2H^2}{8mc^2} \sum_i \langle x_i^2 + y_i^2 \rangle$$

Treating the magnetic field acting on an atom as an external parameter λ of the system, the general relationships in Section 8 can be applied. Namely, if the energy differential is $\Lambda\, d\lambda$ the expression for the mean value is $\bar{\Lambda} = \partial F/\partial\lambda$.

In our case $dE = -(\mathbf{m}\cdot d\mathbf{H})$, where $\mathbf{m}$ is the magnetic moment due to the field. Therefore

$$\overline{\mathbf{m}} = -\frac{\partial F}{\partial \mathbf{H}} = -\frac{e^2\mathbf{H}}{4mc^2} \sum_i \langle x_i^2 + y_i^2 \rangle \tag{32.17}$$

The expression (32.17) does not involve Planck's constant in explicit form, but it should be remembered that any length in atomic units [Sec. 29] is proportional to $h^2/(mc^2)$.

If the system is centrally symmetrical, then

$$\langle x_i^2 \rangle = \langle y_i^2 \rangle = \frac{1}{3} \langle r_i^2 \rangle$$

and

$$\overline{m} = -\frac{e^2H}{6mc^2} \sum_i \langle r_i^2 \rangle \tag{32.18}$$

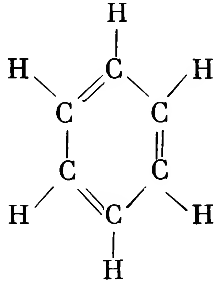
The magnetic polarization M is equal to the density of the atoms or molecules multiplied by $\overline{m}$. Let us evaluate it for the density of a condensed medium being $N \sim 5 \times 10^{22}$ cm^{-3}. Then

$$\frac{Ne^2}{6mc^2} = 2.5 \times 10^9$$

Assuming $\langle r^2 \rangle \sim 10^{-16}$ cm^2 and taking into account that in magnetic polarization the ratio between the magnetic field and magnetic induction involves the factor 4π, we find that the magnetic permeability of diamagnetic substances differs from unity by a quantity of the order 10^{-6}. We could therefore assume that the atoms are in vacuum and not in a field.

The diamagnetic permeability of diamagnetic molecules of cyclic compounds, such as benzene, differs from unity by much more than

that of atoms. A benzene molecule has the following structure:



We see that single and double bonds alternate in the ring. It is proved in the quantum theory of chemical affinity that along such a ring electrons can move freely from atom to atom. In this case r is taken to be the radius of the whole ring and not of a separate atom. This agrees with the high diamagnetic permeability of benzene.

In the case of a saturated compound, cyclohexane C_6H_{12}, where the bonds in the ring are double, the diamagnetic permeability is not of an anomalous value.

Diamagnetism of Free Electrons. It was shown in Exercise 7, [Sec. 14], that electrons in a magnetic field move along helical lines whose axes coincide with the direction of the field. It would appear to follow from this that their motion produces a magnetic moment directed against the external field, so that a gas made up of free electrons would be diamagnetic. But this contradicts the general Van Leeuwen theorem.

The paradox resolves itself in the following way. Statistical equilibrium can be achieved only if a gas is contained in a closed volume (it would be meaningless to calculate the partition function for a nonequilibrium state). Electrons striking the walls restricting the volume bounce back. As a result current is generated by the reflected electrons in the space adjoining the walls; the current produces a magnetic moment directed in the opposite direction of the magnetic moment of the bulk current. It can be shown that in the classical approximation they always cancel out.

In 1930, L. D. Landau observed that compensation occurred only in the purely classical motion of electrons along their paths. Since the displacement of electrons perpendicular to the field is finite [Sec. 5], the quantum theory yields a discrete energy component [Sec. 28]. This means that the classical partition function is in part replaced by a sum, and the Van Leeuwen theorem is inapplicable.

After calculating this sum (first for the case of weak magnetic fields), Landau discovered the *diamagnetic susceptibility* of an elec-

tron gas. Its value we will discuss later. Subsequently it was found that in strong fields the dependence of magnetic susceptibility on the field is of an oscillating nature.

Paramagnetism. The first term in the right-hand side of (32.16) differs from θ if the angular momentum of the system in the ground state is not zero. Let the total angular momentum be J, and let the magnetic field be so weak as not to break the coupling in the L, S-multiplet [Sec. 33]. In such magnetic field there is an anomalous Zeeman effect. The quantum mechanical mean value is given by the expression

$$\langle L_z + 2S_z \rangle = gJ_z$$

where g is the Lande factor [33.51] equal to

$$g = 1 + \frac{J(J+1) + S(S+1) - L(L+1)}{2J(J+1)} \tag{32.19}$$

The free energy term involving J is equal to

$$(\Delta F)_{\text{paramagnetic}} = -\theta \ln \sum_{-J}^{J} e^{-\beta g H J_z/\theta} \tag{32.20}$$

Summing this geometrical progression with respect to J_z and multiplying the numerator and denominator of the sum by $e^{\beta g H/(2\theta)}$, we obtain

$$(\Delta F)_{\text{paramagnetic}} = -\theta \ln \frac{\sinh [\beta g H (J + 1/2)/\theta]}{\sinh [\beta g H/(2\theta)]} \tag{32.21}$$

At sufficiently small values of H the hyperbolic sines expand into series $\sinh x \approx x(1 + x^2/6)$, after which the logarithm of their ratio must also be expanded. Then the free energy component proportional to the square of the magnetic field is equal to

$$(\Delta F)_{\text{paramagnetic}} = -\frac{\beta^2 g^2 H^2}{6\theta} J(J+1) \tag{32.22}$$

From this we obtain the mean component of the magnetic moment parallel to the field:

$$\overline{m} = -\frac{\partial (\Delta F)_{\text{paramagnetic}}}{\partial H} = -\frac{\beta^2 g^2 J(J+1)}{3\theta} H \tag{32.23}$$

which is analogous to the mean component of the electric dipole moment calculated using the Langevin function (Exercise 3, Section 30). The factor of H in the expression for magnetic polarization at room temperature now has a value of the order

$$\frac{N\beta^2 g^2}{3\theta} \approx \frac{5 \times 10^{22} \times 10^{-40}}{3 \times 3 \times 1.4 \times 10^{-14}} \approx 0.5 \times 10^{-4}$$

which is much greater than the magnetic susceptibility of a diamagnetic substance. Therefore at $J \neq 0$ bodies always display para-

magnetic properties compared to which diamagnetism yields only a small correction.

Note that in a weak magnetic field the magnetic polarization due to J is inversely proportional to the absolute temperature.

The paramagnetic properties described here are readily observed in the rare earths and their salts. In these elements the $4f$ shell is being filled, the shell principally lying inside the filled atomic shells [Sec. 33]. Therefore the total angular momentum of the $4f$ shell, which is screened by the outer electrons, freely aligns in space. An external magnetic field acts on it more or less as though it were unaffected by the electric field of the surrounding atoms, and the formulas obtained here hold well for rare earths. In the most general case, when the angular momentum belongs to the outer shells of the atom, crystalline bodies display a dependence of energy on the orientation of the angular momentum with respect to the crystallographic axes. In that case the simple equation (32.20) does not hold.

Paramagnetism of Alkali Metals. Alkali metals display paramagnetism which does not depend on temperature. Pauli offered the following explanation. As mentioned in Section 6, the electrons of alkali metals can be treated as a Fermi gas filling a least-energy sphere in momentum space. Each phase cell is occupied by two electrons. If an external magnetic field is applied, the energies of the electrons whose magnetic moment is directed against the field will become by $2\beta H$ greater than the moment of those parallel to the field. Therefore the least energy of the gas corresponds to a configuration in which the sphere containing electrons with spins parallel to the field is somewhat larger than the sphere with spins antiparallel to the field. This yields a net magnetic polarization of the gas.

The state of an electron Fermi gas in an alkali metal at normal temperature differs little from its state at absolute zero. Therefore, to calculate the main part of the magnetic polarization not dependent on temperature it is sufficient to determine the minimum total, not free, energy of an ideal gas in a magnetic field. At $\theta = 0$ we obtain $F = E - \theta S = E$.

Let us assume that n electrons in a unit volume "flipped" their spins. The corresponding energy of this spin flip is equal to $-2n\beta H$. Now determine the change in the kinetic energy of the gas. The boundary momentum of the electrons whose spin is parallel to the field is equal to

$$p_{0+} = \left(\frac{3}{4\pi}\right)^{1/3} 2\pi h \left(\frac{N}{2} + n\right)^{1/3} \tag{32.24}$$

(see (6.6) and [33.26]).

In the same way we determine p_{0-} for the electrons with opposite spin. Here and further N and n refer to a unit volume. The kinetic energy of a gas under these conditions is

$$E_{\text{kinetic}} = \frac{4\pi}{(2\pi h)^3\, 2m}\left(\int_0^{p_{0+}} p^4\,dp + \int_0^{p_{0-}} p^4\,dp\right)$$

$$= \frac{4\pi}{5}\,\frac{(p_{0+}^5 + p_{0-}^5)}{(2\pi h)^3\, 2m} \tag{32.25}$$

Since in a weak field $n \ll N$, we expand p_{0+}^5 and p_{0-}^5 in a series in the small ratio n/N. Retaining only quadratic terms, we find the required total energy increment:

$$\Delta E = \frac{\pi^{4/3}}{3^{1/3}}\,\frac{h^2 n^2}{mN^{1/3}} - 2n\beta H \tag{32.26}$$

From this minimum condition of ΔE we determine the magnetic polarization M (ΔE is minimal when the total energy gradients of the electrons of both spin orientation are equal):

$$M = 2\beta n = \frac{3^{1/3}}{\pi^{4/3}}\,\frac{\beta^2 m}{h^2}\,N^{1/3} H \tag{32.27}$$

When the diamagnetism of free electrons is taken into account, this quantity must (according to Landau) be reduced by 1/3.

EXERCISES

1. Calculate the dependence of temperature on the magnetic field in isentropic demagnetization of a paramagnetic substance. Neglect the energy transfer from the magnetic system to other degrees of freedom in the medium. Assume the field to be weak.

Solution. From (32.22) the entropy due to the magnetic moments of the substance is

$$\Delta S = -\left(\frac{\partial \Delta F}{\partial \theta}\right)_H = -\frac{\beta^2 g^2 H^2}{6\theta^2} - J(J+1)$$

In isothermal magnetization entropy decreases, since the moments in the field become ordered. In isentropic demagnetization the temperature of the magnetic subsystem in the absence of energy exchange with other degrees of freedom (for example, lattice vibrations) decreases proportional to the magnitude of the field. This is how low temperatures are achieved, if the initial state of the whole system corresponds to low thermal excitation energy—of the order of 1 K. Since the entropy of the lattice at low temperatures is proportional to θ^3 (see Sec. 4), further evening out of the temperature

between the magnetic subsystem and the lattice leaves the overall equilibrium temperature very low.

2. Determine the temperature correction to the magnetic susceptibility of alkali metals, using (6.18).

Hint. It is necessary to go over from the total energy E to the free energy F.

33

FERROMAGNETISM

After H.C. Oersted discovered the magnetic effect of an electric current in 1820, it was A.M. Ampère who suggested that the magnetic properties of iron are due to circular currents flowing within molecules. According to Ampère, magnetization of iron occurs because the electrical moments of the circular currents become parallel, and they are kept in this position by magnetic forces, like compass needles lined up tip to tip.

Ampère's hypothesis seemed virtually obvious until the elementary magnetic moment, that is, the Bohr magneton, and the distance between atoms were established. The interaction energy of two magnets, as of any dipoles [Exercise 2, Section 16], is of the order of the square of the moment divided by the cube of the distance, that is, in this case $10^{-40}/10^{-24} = 10^{-16}$. In thermal units this is 1 K. The thermal motion of atoms would be expected to disturb the order in the configuration of the moments already at 1 K. Yet iron loses its magnetic properties at a temperature around 10^3 K (the *Curie point*). That is why the magnetization ability of iron is not so easily explainable in terms of classical magnetostatics.

The Exchange Energy of a Ferromagnetic. The interaction energy of two elementary magnetic moments is a quantity of the order $\beta^3/a^3 \sim (e^2h^2)/(4a^3m^2c^2)$, that is, it is a relativistic quantity, involving the square of the speed of light in the denominator. The interaction which orders the moments in a ferromagnetic substance is approximately one thousand times greater. It must therefore be of an electrostatic nature and at the same time depend on the spin orientation of individual atoms with respect to one another. Such interaction was examined in [Sec. 34] in connection with the question of the stability of the hydrogen molecule. With antiparallel spins hydrogen molecules attract, with parallel spins they repulse.

Since, in accordance with Pauli's exclusion principle, the wave function of two electrons must be antisymmetric, in the initial approximation it has the form

$$\Psi = \frac{1}{\sqrt{2}} [\psi_a(1)\,\psi_b(2) \mp \psi_a(2)\,\psi_b(1)]$$

The minus corresponds to an antisymmetric spatial wave function and, accordingly, a symmetric spin function, that is, to parallel spins. The plus refers to antiparallel spins. But it is obvious that with the plus sign a spatial wave function corresponds to a smaller energy of the system, since in that case a symmetric wave function does not have nodal surfaces anywhere, which means that in the ground state both ψ_a and ψ_b do not have them either. An antisymmetric spatial wave function vanishes on the median plane between nuclei a and b.

Therefore a hydrogen molecule can be stable only if the spins are antiparallel. The $3d$ shell of the ferromagnetic elements Fe, Co, Ni is not filled. It has been suggested that for them the parallel spin configuration is stable. This is possible because the wave functions of $3d$ states have nodal surfaces themselves. That is why, unlike the $1s$ states of hydrogen, the symmetry of the general wave function of two electrons does not predetermine a more stable state.

In view of the complexity of quantum mechanical calculations of many-electron systems, this has not as yet been confirmed by direct computations, although it seems highly probable.

Since the $3d$ shell of atoms of ferromagnetic substances lies inside the atom, the interaction energy difference for parallel and antiparallel spins is much less than in the case of a hydrogen molecule. The value of the integral is also affected by the fact that the wave functions of individual electrons have nodal surfaces. Hence, the absolute value of the exchange integral (see [34.8] and (33.20])

$$e^2 \int \psi_a(1)\,\psi_b(2)\,\frac{1}{r_{12}}\,\psi_a(2)\,\psi_b(1)\,dV_1\,dV_2$$

is much less for ferromagnetic substances than for the hydrogen molecule. It can be seen from the Curie point that the value of the exchange integral for iron is a quantity of the order of 0.1 eV (10^3 K).

Exchange energy depends only on the mutual spin orientation and is not dependent on the spin orientation with respect to the crystalline lattice. Therefore, the nonrelativistic exchange energy of ferromagnetic substances is associated only with the absolute value of the resultant moment, or the magnetic polarization M, and not with the direction of $\mathbf{M}$ in the crystal.

Ferromagnetic substances resemble pyroelectric crystals (Sec. 30), which have a total electric polarization, $\mathbf{P}$. Unlike ferromagnetic substances, in the latter not only the value of $\mathbf{P}$ but its direction as well is determined by electrostatic forces in the lattice. Therefore

pyroelectricity can exist only in low-symmetry crystals without defined axes or with only one. Ferromagnetism does not require low lattice symmetry. Fe, Co and Ni crystals have cubic symmetry. Magnetization along one of the axes has a very small effect (in relativistic order of magnitude) on the cubic state of these crystals.

The Curie Point. We shall now consider the properties of ferromagnetic substances close to the Curie point. At the Curie point M becomes zero, not abruptly, but gradually approaching zero as the temperature rises. That was the first investigated phase transition of the second kind (Sec. 11). Since near the Curie point the magnetic polarization M is small, the thermodynamic potential in its neighbourhood can be expanded in a power series in M^2. The components of **M** cannot be involved individually since the exchange energy depends only on the absolute value M. Furthermore, it is assumed that the derivatives of the thermodynamic potential with respect to these components do not become infinite at $M = 0$. That is why the expansion is in powers of M^2

Suppose there is no magnetic field. Then the expansion of the thermodynamic potential up to terms quadratic with respect to M^2 is of the form

$$f'_M = aM^2 + \frac{b}{2} M^4 \tag{33.1}$$

Here the density of the thermodynamic potential is designated by the same letter as the free energy density, but in this case it does not lead to confusion.

For there to exist an equilibrium magnetization with a finite value of M the factor b in front of M^4 must be positive; otherwise f'_M would have no minimum. We represent the factor a as

$$a = \alpha(\theta - \theta_c) \tag{33.2}$$

where θ_c is the *Curie temperature*. Then from the condition $\partial f'_M/\partial M^2 = 0$ we obtain

$$M^2 = \frac{\alpha}{b}(\theta_c - \theta) \tag{33.3}$$

Consequently, the magnetization really does become zero at the Curie temperature. Obviously, if $\alpha > 0$ then the real magnetization values are achieved only at temperatures below the Curie point, $\theta < \theta_c$. Here polarization M is proportional to $(\theta_c - \theta)^{1/2}$, which satisfactorily agrees with experience when θ is close (but not too close) to θ_c. This proves that the assumption concerning the form of the dependence of the expansion coefficient a upon the temperature in this temperature region is valid. In the most general case we can only require that the condition $a(\theta_c) = 0$ be satisfied.

In the quantum theory, the partition functions defining the thermodynamic potential have practically infinite multiplicity for large systems of interacting particles. Therefore, the dependence of such functions on the parameters involved in them may not be analytical, that is, they may not allow for a series expansions of the type (33.1) close to transition points. Experience nevertheless confirms the analytical form of the dependence of the thermodynamic potential on magnetic polarization in a temperatures range not too close to the Curie point.

The dependence of entropy on M^2 is given by the derivative

$$S_{\mathrm{m}} = -\left(\frac{\partial f'_M}{\partial \theta}\right)_M = -\frac{\partial}{\partial \theta}\,\alpha(\theta-\theta_c)\,M^2 = -\alpha M^2 \qquad (33.4)$$

Here the term $bM^4/4$ need not be taken into account since at small M^2 its contribution is negligible. Substituting (33.3), we obtain

$$\begin{aligned} S_{\mathrm{m}} &= -\frac{\alpha^2}{b}(\theta_c-\theta) \quad &&\text{at} \quad \theta<\theta_c \\ &= 0 &&\text{at} \quad \theta>\theta_c \end{aligned} \qquad (33.5)$$

It can be seen from this that at the Curie point the specific heat $\theta(\partial S_{\mathrm{m}}/\partial\theta)_M$ experiences a discontinuity

$$\Delta c_p = \theta\left(\frac{\partial S_{\mathrm{m}}}{\partial \theta}\right)_M = \frac{\alpha^2\theta_c}{b} \qquad (33.6)$$

This agrees with the general theory of phase transitions of the second kind (Sec. 11). The specific heat of the magnetically disordered phase is greater because it possesses the entropy of random magnetic moment alignment.

In a magnetic field **H** expression (33.1) receives an increment $-(\mathbf{M}\cdot\mathbf{H})$. The minimum condition for f'_M is then of the form

$$\frac{\partial f'_M}{\partial \mathbf{M}} = 2\mathbf{M}\,\{\alpha(\theta-\theta_c)+bM^2\} - \mathbf{H} = 0 \qquad (33.7)$$

From this we define the coefficient of proportionality between field and polarization at temperatures above the Curie point, when the polarization in the absence of a field, M, is equal to zero:

$$\mathbf{M} = \frac{1}{2\alpha(\theta-\theta_c)}\,\mathbf{H} \qquad (33.8)$$

This resembles the dependence (32.23) for paramagnetic substances, but instead of the temperature θ the denominator contains the difference $\theta-\theta_c$ (the *Curie-Weiss law*).

Energy of Magnetic Anisotropy. The dependence of energy on the direction of magnetic polarization in a lattice is due to the spin-orbital interaction and is several orders of magnitude (three or four)

smaller than the exchange energy [Sec. 33]. The connection between spin and the orbital motion of electrons is of a relativistic nature and is therefore weaker than the electrostatic exchange interaction.

Taking into account that the value of those energy terms that describe its dependence on orientation in a crystal are small, we find that it is sufficient to retain the first nonzero terms in the expansion. They correspond to quantum-mechanical perturbation theory expansions of the same order, which is why the question of the possibility of expanding the function into a series does not arise here. The type of expansion must be compatible with the symmetry of the crystal (as was the case with crystalline dielectrics, Sec. 30). The obtained expression is conventionally called the *energy of magnetic anisotropy*.

This expression must be an even function of the magnetic polarization components, since the energy does not change its sign under time inversion $t \to -t$, while vector $\mathbf{M}$ transforms into $-\mathbf{M}$. The simplest even function of $\mathbf{M}$ is the quadratic form with respect to the projections M_x, M_y, M_z. The coefficients of this form make up a symmetric tensor of rank 2. As in the case of the dielectric permittivity tensor, the number of independent components of this tensor is determined by the point symmetry of the crystal. If it has no axes of symmetry of order n greater than two, the tensor has three independent eigenvalues. This case will not be considered since it is not typical of ferromagnetic substances.

If there is a three-fold, four-fold, or six-fold axis, two eigenvalues in a plane perpendicular to it are the same. In crystals with cubic symmetry the tensor of rank 2 degenerates into a scalar, since it has three identical eigenvalues.

Thus, given two different eigenvalues of the tensor, the magnetic anisotropy energy can be written as follows:

$$f_a' = \frac{\beta_1}{2}(M_x^2 + M_y^2) + \frac{\beta_2}{2} M_z^2$$

$$= \frac{\beta_1}{2}(M_x^2 + M_y^2 + M_z^2) + \frac{\beta_2 - \beta_1}{2} M_z^2$$

But the first term involves no anisotropy at all. It can be considered the isotropic exchange energy or neglected altogether. Instead of M_z^2 we write $M^2 \cos^2 \theta$ (where θ is the angle between the polarization vector and the symmetry axis). Substituting β for the coefficient $\beta_2 - \beta_1$, we obtain

$$f_a' = \frac{\beta}{2} M^2 \cos^2 \theta \tag{33.9}$$

If $\beta < 0$, the anisotropy energy has the least, that is, the equilibrium, value at $\theta = 0$. The polarization vector is directed along the symmetry axis, which in this case is called the *axis of easy mag-*

netization. At $\beta > 0$ equilibrium corresponds to $\theta = \pi/2$. The crystal is magnetized in the x,y-plane, but to determine the direction of **M** in that plane terms of powers higher than the second must be taken into account in the magnetic anisotropy energy.

In a cubic crystal, such as Fe, Co or Ni, the tensor β has one independent component. The quadratic form degenerates into a scalar expression involving only M^2. The expansion of f'_a involves only even functions of **M**, so that now it is necessary to refer to fourth-order terms. To possess cubic symmetry they must not change in a rearrangement of the x, y, and z coordinates. Therefore

$$f'_a = \beta'_1 (M_x^4 + M_y^4 + M_z^4) + \beta'_2 (M_x^2 M_y^2 + M_x^2 M_z^2 + M_y^2 M_z^2)$$

$$= \left(\beta'_1 - \frac{\beta'_2}{2}\right) (M_x^4 + M_y^4 + M_z^4)$$

$$+ \frac{\beta'_2}{2} (M_x^4 + M_y^4 + M_z^4 + 2M_x^2 M_y^2 + 2M_x^2 M_z^2 + 2M_y^2 M_z^2)$$

$$= \left(\beta'_1 - \frac{\beta'_2}{2}\right) (M_x^4 + M_y^4 + M_z^4) + \frac{\beta'_2}{2} M^4$$

The last term does not involve anisotropy, so that f'_a reduces to the form

$$f'_a = \beta' (M_x^4 + M_y^4 + M_z^4) \tag{33.10}$$

Let us determine the extremum of this expression for a given value of M. The sum of the squares of the three direction cosines is unity. If one of them is unity, the other two are zero. This corresponds to **M** directed along a side of the cube. If two cosines are equal to $1/\sqrt{2}$ and the third is zero, **M** is directed along a diagonal of a face of the cube. Finally, if all three cosines are equal to $1/\sqrt{3}$, **M** is directed along a spatial diagonal of the cube. For these cases the sum of the fourth powers of the cosines are equal to 1, 1/2, and 1/9. Hence, if $\beta' < 0$, the directions of easy magnetization coincide with the sides of the cube, that is, with four-fold axes. This is the case of iron. If $\beta' > 0$, the spatial diagonals of the cube become the easy magnetization directions. This is the case with cobalt. As mentioned before, symmetry in a magnetized cubic crystal is only slightly disturbed, because f'_a is of relativistic order of magnitude.

A Crystal in a Magnetic Field. We shall consider only a crystal with one axis of easy magnetization. As stated before, in a magnetic field the term $-(\mathbf{M}\cdot\mathbf{H})$ is added to the expression for the free energy. Suppose that $\beta < 0$. Let us determine the direction of vector **M**.

We take the x axis in a plane through the easy magnetization axis z and the magnetic field. This in no way restricts the general case. Then $M_z = M\cos\theta$, $M_x = M\sin\theta$. The scalar product $(\mathbf{M}\cdot\mathbf{H})$ takes the form $M(H_x \sin\theta + H_z \cos\theta)$. To determine θ, that is,

the direction of the polarization vector in the magnetic field, we must find the minimum of

$$f'_a - (\mathbf{M}\cdot\mathbf{H}) = -\frac{|\beta|}{2} M^2 \cos^2\theta - M(H_x \sin\theta + H_z \cos\theta) \tag{33.11}$$

Differentiating with respect to θ, we arrive at the equation

$$|\beta| M^2 \sin\theta\cos\theta - MH_x \cos\theta + MH_z \sin\theta = 0 \tag{33.12}$$

We introduce the notation $\cos\theta \equiv \xi$ and then divide (33.12) by $\sin\theta\cos\theta$ to get

$$|\beta| M^2 - \frac{MH_x}{(1-\xi^2)^{1/2}} + \frac{MH_z}{\xi} = 0$$

To get rid of the irrationality, we transfer the second term to the right and square:

$$\left(|\beta| M^2 + \frac{H_z}{\xi}\right)^2 = \frac{H_x^2}{1-\xi^2} \tag{33.13}$$

Reducing this equation to a common denominator yields an equation of the fourth power. Its real roots lie at $|\xi| < 1$, because the left-hand side is always positive. An equation with real coefficients can have only pairs of conjugate complex roots. Therefore there are either two or four real roots.

Since the maxima and minima of function (33.11) alternate, it has either one or two minima. One of the two is perfectly stable, the other is metastable. There are two equilibrium positions, because at $H = 0$ the directions $\theta = 0$ and $\theta = \pi$ are equivalent. In a weak field the polarization partially deviates from these directions, while in a strong field they converge into one. The convergence boundary will be determined in Exercise 1.

Thanks to the two equilibrium directions of $\mathbf{M}$, the sequence of states through which a crystal passes in magnetization and demagnetization may vary, provided the processes are not infinitely slow. The metastable state persists for a long time and does not have time to develop into total stability. But this means that the process is irreversible, since it does not pass through a sequence of perfectly equilibrium states [Sec. 8]. The magnetization curve does not follow the demagnetization curve. This phenomenon is called *hysteresis*. It should be distinguished from hysteresis phenomena in the magnetization of polycrystalline samples. The latter is of great importance in electrical engineering, but there hysteresis is due to magnetic forces between the crystals and magnetoelastic forces.

If the magnetic field is perpendicular to the easy-magnetization axis, that is, $H_z = 0$, the minimum condition takes the form

$$\sin\theta = \frac{H}{|\beta| M} \tag{33.14}$$

When this ratio is less than unity, the equation has two roots, θ_m and $\pi - \theta_m$. But if the whole crystal is magnetized at one angle, say θ_m, it produces an external magnetic field whose energy is always positive. Its energy is added to the free energy of the crystal, which becomes far removed from the minimum necessary for equilibrium. Closer to equilibrium there is a configuration in which the polarization vector in the crystal layers alternately forms angles θ_m and $\pi - \theta_m$ with the easy-magnetization axis. These layers are called *domains*. The resultant external field of the crystal is thereby weakened, which leads to a decrease in the total energy.

Domains. Consider a ferromagnetie crystal with one direction of easy magnetization, in the absence of an externally applied magnetic field. If, as a whole, the crystal has one value of the magnetic polarization **M**, there will appear, as was just mentioned, an external field of its own. The volume energy of such a field increases as the integral of the square of the field with respect to the volume, that is, as the cube of the dimensions of the crystal. From the energy point of view it is therefore more advantageous for a monocrystal to divide into domains magnetized in opposite directions. This weakens the external field, but at the same time there appears the energy of the intermediate transition zones between the domains, where the polarization necessarily is at an angle to the axis of easy magnetization.

The actual separation into domains takes place according to the minimum of the total energy comprising the energy of the external magnetic field and the total energy of the transition layers. They are both positive, and the minimum of their sum, which depends on the size and form of the monocrystal, can be found. L. D. Landau and E. M. Lifshits showed how to determine the structure and energy of a transitional layer. In such a layer the polarization vector between uniformly magnetized domains gradually turns through 180°. Consequently, it has a variable spatial direction. This leads to an increase in the exchange energy of a ferromagnetic substance, because the least energy of this nature corresponds to unidirectional moments.

If the domains come to the edge of the monocrystal with no change in direction of polarization, then magnetic field lines emerge from them into the surrounding medium. By virtue of the boundary condition (28.33), in this case $\mathbf{B} = \mathbf{H}$, while in a ferromagnetic substance $\mathbf{B} = 4\pi\mathbf{M}$, since there $\mathbf{H} = 0$. Hence, if the polarization of the domain is perpendicular to the surface of the monocrystal, then $H_n^{\text{ext}} = 4\pi M$. But the external field possesses additional energy. Therefore if the anisotropy energy is not too great, the configuration shown in Figure 45 is more favourable. The arrows correspond to the polarization directions in the domains. As can be seen from the

drawing, the vector lines are closed, so that the equation div $\mathbf{M} = 0$ is satisfied.

In small domains of triangular cross section magnetic polarization does not coincide with the axis of easy magnetization, and such domains increase the free energy of the monocrystal. But on the other hand, the magnetic field does not escape outside, since the polarization vector is everywhere tangent to the surface. Thus, the appearance of small domains at the surface depends on what

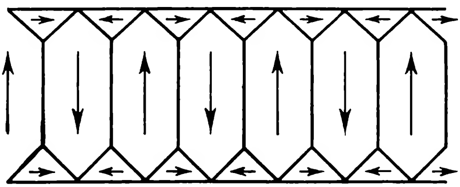


Figure 45

is favoured from the energy point of view: emergence of the magnetic field outside the monocrystal, or magnetization perpendicular to the axis of easy magnetization.

This configuration was predicted theoretically and subsequently discovered experimentally. The surface of a ferromagnetic crystal was coated parallel to the polarization vector with an emulsion of very light colloid ferromagnetic particles. The particles gathered at the domain boundaries, as shown in Figure 45. Attraction to the boundaries is due to the appearance near them of microirregularities in the magnetic field associated with the fact that the direction of $\mathbf{M}$ changes. Magnetic substances, as we know from [17.35], are drawn into the nonuniform field region.

Very small ferromagnetic crystals consist of one domain, because the volume energy of the magnetic field is in this case small (proportional to the cube of the dimensions). The interface energy between domains is proportional to the square of the dimensions (the area of the surface). Obviously, for very small dimensions uniform magnetization is favoured from the energy point of view, while for large dimensions division into domains is favoured.

Antiferromagnetism. If in some crystalline medium the sign of the exchange integral is opposite to the sign that should be in a ferromagnetic, another type of ordering of the magnetic moments is possible: the moments of neighbouring atoms are oppositely oriented. At some temperatures this order may disappear, but without any significant restructuring of the lattice.

This is also a Curie point, in which a discontinuity in the specific heat occurs, but without any apparent change in the magnetic properties of the substance. The discontinuity appears inevitably, owing to the change in the temperature dependence of the entropy in phases with various degrees of ordering.

This type of phase transitions of the second kind was pointed out by Landau in analyzing some experimentally observed discontinuities in specific heat. Subsequent experiments in magnetic neutron scattering confirmed that the symmetry of a crystal's magnetic properties really does change at the transition point. As long as the spins of neighbouring atoms are oppositely oriented, the magnetic spacing of the lattice is over every other atom. When the directions of all spins are equiprobable, the magnetic spacing is the same as the structural spacing. This is observed in neutron diffraction due to their magnetic interactions with atomic spins.

Owing to spin-orbital interaction, the magnetic moments of neighbouring atoms in an *antiferromagnetic* phase may not cancel out completely. This happens when the positions of identical atoms possessing magnetic moments in the lattice are not equivalent and the electric field near them is different. In that case the interaction of the atomic spins with electron orbital momenta results in the total spin of neighbouring atoms not being strictly equal to zero (as required by exchange forces) and having a small net value. The smallness of this quantity is determined by the ratio of the spin-orbital forces to the exchange forces. A crystal with such weak ferromagnetism is called *ferrimagnetic*, because the property is frequently observed in iron compounds.

EXERCISE

Determine in coordinates H_x, H_z the boundary of the region in which a ferromagnetic crystal with one easy-magnetization axis has two equilibrium polarization directions, one stable and one metastable.

Solution. The boundary of the region corresponds to the merging of the roots of Eq. (33.12):

$$\frac{H_x}{\sin\theta}-\frac{H_z}{\cos\theta}-|\beta|M=0$$

The roots merge when they become common to this equation and the equation obtained as a result of differentiating it, that is,

$$\frac{H_x}{\sin^3\theta}+\frac{H_z}{\cos^3\theta}=0$$

From this equation,

$$\tan^3\theta=-\frac{H_x}{H_z}$$

Substituting into the initial equation, we find the relation between H_x and H_z that defines the required boundary:

$$H_x^{2/3} + H_z^{2/3} = (|\beta| M)^{2/3}$$

This is a closed, curvilinear, starlike figure with the points on the axes (an astroid). The metastable region lies inside it.

34

THE MAGNETIC FIELD OF DIRECT CURRENT

Basic Equations. In Section 31 we obtained an equation describing the magnetic field produced by direct current through conductors:

$$\operatorname{curl} \mathbf{H} = \frac{4\pi}{c} \mathbf{j} \tag{34.1}$$

where $\mathbf{j}$ is the current density. Furthermore, from (28.5),

$$\operatorname{div} \mathbf{B} = 0 \tag{34.2}$$

If $\mathbf{B}$ and $\mathbf{H}$ are connected by a proportional dependence, that is,

$$\mathbf{B} = \mu \mathbf{H} \tag{34.3}$$

then in a medium with constant permeability

$$\operatorname{curl} \mathbf{B} = \frac{4\pi}{c} \mu \mathbf{j} \tag{34.4}$$

To satisfy Eq. (34.2), we assume, as usual, that

$$\mathbf{B} = \operatorname{curl} \mathbf{A} \tag{34.5}$$

where $\mathbf{A}$ is the vector potential of the magnetic field. It is convenient to impose a condition similar to the Lorentz condition [17.7]:

$$\operatorname{div} \mathbf{A} = 0 \tag{34.6}$$

because vector $\mathbf{A}$ cannot be defined only by its curl. The condition (34.6) can always be satisfied by adding to the vector potential a gradient of some scalar quantity, which does not affect $\mathbf{B}$. Substituting (34.5) into (34.4), we obtain

$$\operatorname{curl} \operatorname{curl} \mathbf{A} \equiv \operatorname{grad} \operatorname{div} \mathbf{A} - \nabla^2 \mathbf{A} = \frac{4\pi}{c} \mu \mathbf{j} \tag{34.7}$$

or, from condition (34.6),

$$\nabla^2 \mathbf{A} = -\frac{4\pi}{c}\mu\mathbf{j} \tag{34.8}$$

If the magnetic permeability is not constant everywhere but only in individual spatial regions, then the usual conditions (28.33) and (28.38) are satisfied on the boundaries where μ experiences a discontinuity. Written in terms of the vector potential they have the form

$$\operatorname{curl}_n \mathbf{A}_1 = \operatorname{curl}_n \mathbf{A}_2, \qquad \frac{1}{\mu_1}\operatorname{curl}_t \mathbf{A}_1 = \frac{1}{\mu_2}\operatorname{curl}_t \mathbf{A}_2 \tag{34.9}$$

where n and t as usual denote the normal and tangential components.

The Vector Potential and Field in an Unbounded Medium. If a medium is unbounded and homogeneous, that is, the magnetic permeability has the same value everywhere, the vector potential for a given current distribution is

$$\mathbf{A} = \frac{\mu}{c}\int \frac{\mathbf{j}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}\,dV' \tag{34.10}$$

This is a solution of the Laplace equation with a right-hand side, a solution analogous to [17.11]. The integral is taken over the whole current distribution $\mathbf{j}(\mathbf{r}')$. Hence the magnetic induction is

$$\mathbf{B} = \operatorname{curl}\mathbf{A} = \frac{\mu}{c}\int \left(\operatorname{grad}\frac{1}{|\mathbf{r}-\mathbf{r}'|}\times \mathbf{j}(\mathbf{r}')\right) dV' \tag{34.11}$$

Since the differentiation is carried out with respect to the coordinates of the point at which the induction is observed, and the integration is with respect to the argument $\mathbf{r}'$, we can first find $\operatorname{grad}(|\mathbf{r}-\mathbf{r}'|)^{-1}$. Differentiating and rearranging the factors in the vector product, we obtain

$$\mathbf{B} = \frac{\mu}{c}\int \frac{dV'}{|\mathbf{r}-\mathbf{r}'|^3}\,\mathbf{j}\times(\mathbf{r}-\mathbf{r}') \tag{34.12}$$

One has to deal most commonly with thin linear conductors. The current direction is determined according to a linear element of the conductor. Therefore

$$\mathbf{j}\,dV' = I\,d\mathbf{l} \tag{34.13}$$

where I is the total current in the conductor. The transition (34.13) can be carried out in all cases when the integral does not, as a consequence, diverge (see further on self-inductance). Using (34.13), we write the expression for the vector potential and the magnetic

induction produced by a linear current:

$$\mathbf{A}=\frac{I\mu}{c}\int\frac{d\mathbf{l}}{|\mathbf{r}-\mathbf{r}'|} \tag{34.14}$$

$$\mathbf{B}=\frac{I\mu}{c}\int\frac{1}{|\mathbf{r}-\mathbf{r}'|^3}\,d\mathbf{l}\times(\mathbf{r}-\mathbf{r}') \tag{34.15}$$

(the *Biot-Savart law*).

Field at Great Distances from a System of Currents. Suppose the dimensions of a current-carrying circuit are very small in comparison with the distance to the point at which the vector potential is being determined (see [Sec. 17]). Let us then replace $|\mathbf{r}-\mathbf{r}'|^{-1}$ under the integral sign in (34.14) by its series expansion in powers of $\mathbf{r}'$:

$$\frac{1}{\|\mathbf{r}-\mathbf{r}'|}=\frac{1}{r}+\frac{(\mathbf{r}'\cdot\mathbf{r})}{r^3} \tag{34.16}$$

employed many times in Volume 1. Next r^{-1} and r^{-3} are taken out of the integral sign. Hence

$$\mathbf{A}=\frac{I\mu}{c}\left(\frac{1}{r}\int d\mathbf{l}+\frac{1}{r^3}\int(\mathbf{r}\cdot\mathbf{r}')\,d\mathbf{l}\right) \tag{34.17}$$

But the circuit is assumed closed, and $\int d\mathbf{l}=0$. For this we go over to a scalar element of length according to the formula

$$d\mathbf{l}=\frac{d\mathbf{r}'}{dl}\,dl$$

because $d\mathbf{r}'\equiv d\mathbf{l}$. We use dl as a separate designation only to note integration along a linear conductor. Carrying out the replacement and integrating by parts, we obtain

$$\int(\mathbf{r}\cdot\mathbf{r}')\frac{d\mathbf{r}'}{dl}\,dl=\int dl\,\frac{d}{dl}\,[(\mathbf{r}\cdot\mathbf{r}')\,\mathbf{r}']-\int\left(\mathbf{r}\cdot\frac{d\mathbf{r}'}{dl}\right)\mathbf{r}'\,dl$$

But taken over a closed circuit, the first integral in the right-hand side vanishes because the derivative with respect to l is of a single-valued function. In the second integral we return to $d\mathbf{l}$:

$$\int\left(\mathbf{r}\cdot\frac{d\mathbf{r}'}{dl}\right)\mathbf{r}'\,dl=\int(\mathbf{r}\cdot d\mathbf{r}')\,\mathbf{r}'\equiv\int(\mathbf{r}\cdot d\mathbf{l})\,\mathbf{r}' \tag{34.18}$$

The obtained integral can be replaced by the half-sum of the right- and left-hand sides of Eq. (34.18), then applying the formula for a double vector product:

$$\mathbf{A}=\frac{I\mu}{2cr^3}\int[(\mathbf{r}\cdot\mathbf{r}')\,d\mathbf{l}-(\mathbf{r}\cdot d\mathbf{l})\,\mathbf{r}']=\frac{I\mu}{2cr^3}\,\mathbf{r}\times\int(d\mathbf{l}\times\mathbf{r}) \tag{34.19}$$

Let us now introduce the magnetic moment of the circuit according to the formula

$$\mathbf{m} \equiv \frac{I}{2c} \int \mathbf{r} \times d\mathbf{l} \tag{34.20}$$

(We denote it m instead of $\boldsymbol{\mu}$ to avoid confusion with magnetic permeability μ). The vector potential in this case is expressed as follows:

$$\mathbf{A} = \frac{\mu}{r^3} \mathbf{m} \times \mathbf{r} \tag{34.21}$$

Note that $(1/2)\, \mathbf{r} \times d\mathbf{l}$ is an area element $d\mathbf{S}$ of the circuit. This relationship is used when the law of conservation of momentum is represented as an integral of areas [Sec. 5]. Accordingly, the magnetic moment is

$$\mathbf{m} = \frac{I}{c} \int d\mathbf{S} \tag{34.22}$$

If case of a plane circuit $\int d\mathbf{S}$ must be replaced by an area vector normal to the surface.

The magnetic induction of the circuit is (see [17.21])

$$\mathbf{B} = \frac{3\mathbf{r}\,(\mathbf{m}\cdot\mathbf{r}) - \mathbf{m}r^2}{r^5} \tag{34.23}$$

Work Done by a Magnetic Field on Current. The expression for the work done by a magnetic field (32.4) is transformed in such a way as to involve the current explicitly. For this it is sufficient to replace the magnetic induction **B** by the vector potential according to formula (34.5):

$$dA = \frac{1}{4\pi} \int (\mathbf{H} \cdot d \operatorname{curl} \mathbf{A})\, dV = \frac{1}{4\pi} \int (\mathbf{H} \cdot \operatorname{curl} d\mathbf{A})\, dV \tag{34.24}$$

The obtained integral is transformed by parts in the usual way:

$$dA = \frac{1}{4\pi} \int (\mathbf{H} \cdot \operatorname{curl} d\mathbf{A})\, dV$$
$$= \frac{1}{4\pi} \int \mathbf{H}\, (d\mathbf{S} \times d\mathbf{A}) + \frac{1}{4\pi} \int (d\mathbf{A} \times \nabla)\, \mathbf{H}\, dV$$

The first integral vanishes on a surface where there is no field, that is, sufficiently far away from the system of currents; the second integral, after a cyclic permutation of the factors in the integrand, takes the form

$$dA = \frac{1}{4\pi} \int (d\mathbf{A} \cdot \operatorname{curl} \mathbf{H})\, dV = \frac{1}{c} \int (d\mathbf{A} \cdot \mathbf{j})\, dV \tag{34.25}$$

Equation (34.25) may cause some wonder. We know from Volume 1 that a stationary magnetic field does not do work on charges.

But when a magnetic field varies, an induced electric field appears according to (28.4), which actually does work.

If only the linear dependencies (34.1)-(34.3) are involved in a system of currents, Eq. (34.25) can be integrated with respect to **A**, taking into account that the current density **j** is proportional to the vector potential. We assume that the field is not external and is due to the current **j** itself. Then the free energy of a system of currents is

$$F_{\mathrm{m}} = \frac{1}{2c} \int (\mathbf{A} \cdot \mathbf{j}) \, dV \tag{34.26}$$

We assume the system to be in equilibrium in the first approximation, that is, we disregard losses due to Joule heat. When conductance is high, such an approach is justified, if irreversible evolution of heat during the time the magnetic field changes is not great. The subscript "m" in the free energy indicates that only that part of it must be taken which is due to the current or to the magnetic field.

Substituting the vector potential according to (34.14), we represent F_{m} in the form of a double space integral:

$$F_{\mathrm{m}} = \frac{\mu}{2c^2} \int \int \frac{\mathbf{j}(\mathbf{r})\,\mathbf{j}(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \, dV \, dV' \tag{34.27a}$$

For a system of linear conductors we substitute the current density according to (34.13). Then

$$F_{\mathrm{m}} = \frac{\mu}{2c^2} \sum_{i,\,k} I_i I_k \int \int \frac{d\mathbf{l}_i \, d\mathbf{l}_k}{|\mathbf{r}_i - \mathbf{r}'_k|} \tag{34.27b}$$

Here the summation is carried out along all the separate circuits. The integrals

$$\mathscr{L}_{ik} \equiv \frac{\mu}{c^2} \int \int \frac{d\mathbf{l}_i \, d\mathbf{l}_k}{|\mathbf{r}_i - \mathbf{r}'_k|} \tag{34.28}$$

where $\mathscr{L}_{ik} = \mathscr{L}_{ki}$ at $i \neq k$ is called the *mutual inductance* of the two circuits.

The analogous expression at $i = k$ involves a logarithmic singularity in the integral. However, since a conductor is always of finite thickness, such an integral can always be given approximate meaning (see below). It is then called the *self-inductance* of the ith circuit.

Assuming that this has been done, we obtain the expression for F_{m} as a quadratic form of the currents:

$$F_{\mathrm{m}} = \frac{1}{2} \sum_{i,\,k} \mathscr{L}_{ik} I_i I_k \tag{34.29}$$

Here the $\mathscr{L}_{ik}$'s are given in electromagnetic units so as to eliminate the square of the speed of light c^2 from the denominators.

Let us now obtain another important expression for F_{m}. For linear currents Eq. (34.26) takes the form

$$F_{\text{m}} = \frac{1}{2c} \sum_i I_i \int (\mathbf{A}_i \cdot d\mathbf{l}_i) \tag{34.30}$$

Transforming the line integral according to Stokes' theorem, we obtain

$$F_{\text{m}} = \frac{1}{2c} \sum_i I_i \int (\text{curl}\,\mathbf{A}_i \cdot d\mathbf{S}_i) = \frac{1}{2c} \sum_i \int (\mathbf{B}_i \cdot d\mathbf{S}_i)$$

The magnetic flux linked by the circuit, appearing in this equation, is conventionally denoted Φ_i:

$$\Phi_i \equiv \int (\mathbf{B}_i \cdot d\mathbf{S}_i) \tag{34.31}$$

Thus, the free energy has the form

$$F_{\text{m}} = \frac{1}{2c} \sum_i \Phi_i I_i \tag{34.32}$$

The diverging integrals mentioned before are not involved here explicitly. Comparing formulas (34.32) and (34.29), we obtain the expression for the magnetic flux linked by the ith circuit:

$$\Phi_i = c \sum_k \mathscr{L}_{ik} I_k \tag{34.33}$$

Self-Inductance. The singularity in the double integral (34.28) at $i = k$ is due to the fact that in this case $\mathbf{r}_i - \mathbf{r}'_i$ inevitably vanishes when one and the same circuit is traversed (when the $\mathbf{r}_i$th point coincides with the $\mathbf{r}'_i$th). Omitting the subscripts, we can write the expression for self-inductance of a separate conductor as

$$\mathscr{L} = \frac{\mu}{c^2} \int d\mathbf{l} \int \frac{d\mathbf{l}'}{|\mathbf{r} - \mathbf{r}'|} \tag{34.34}$$

Let us take, for example, a rectilinear wire of length l and radius a. Denoting the distance from one end of the wire by x, we rewrite the expression for self-inductance as follows:

$$\mathscr{L} = \frac{\mu}{c^2} \int_0^l dx \left(\int_0^x \frac{dx'}{x - x'} + \int_x^l \frac{dx'}{x' - x} \right)$$

$$= \frac{\mu}{c^2} \int_0^l dx \left(-\ln (x - x') \Big|_0^x + \ln (x' - x) \Big|_x^l \right) \tag{34.35}$$

The limits of the internal integral cannot be directly substituted into this expression. But the initial formula (34.34) does not hold if the distance from x to x' is of the same order as the diameter of the conductor or less. Therefore the limits of the change in x in the integrals must be replaced by $x - a/2$ and $x + a/2$ (logarithmic precision). The error is the less the greater $\ln(2l/a)$ as compared with unity. After substituting this limit we obtain

$$\mathcal{L} = \frac{\mu}{c^2}\int_0^l dx\left(-\ln\frac{a}{2x} + \ln\frac{2(l-x)}{a}\right) = \frac{2\mu}{c^2}\, l \times \ln\frac{2l}{ea} \quad (34.36)$$

The number $e/2 = 1.359$ under the logarithm sign has been written only so as not to violate mathematical equality. Actually, Eq. (34.36) does not require such precision. Its meaning is that the obtained expression does not depend on the current distribution over the cross section of a thin conductor, which is the same as in the case of mutual inductance. Thereby formula (34.29) is justified (at least as an approximate one).

Self-inductances of coils and solenoids are applied most frequently. Then the approximation just made is replaced by another one. Suppose a cylindrical coil has n winding loops per unit length along the axis. If the winding is thin, the flow of current through the solenoid is equivalent to flow along its surface perpendicular to the generatrix of the cylinder. If the current in the coil is I, the surface current density is $j_t = nI$. In a sufficiently long solenoid the magnetic field inside is much stronger than outside. Therefore in Eq. (28.34) we can neglect the external magnetic field. Then the field within the solenoid can be taken as

$$H = \frac{4\pi nI}{c} \quad (34.37)$$

Denoting the radius of the solenoid as r and the total length as l, we find that the magnetic flux through one turn of the coil is

$$\pi r^2 \mu H$$

and the flux Φ through all nl turns is $\pi r^2 H\mu nl$. In the present case $\mathcal{L} = \Phi/(cI)$, or

$$\mathcal{L} = \frac{4\pi^2\mu n^2 r^2 l}{c^2} \quad (34.38)$$

This expression is the more exact the greater the ratio of the length of the solenoid to its radius and the thinner the winding.

In "soft" polycrystalline ferromagnetic materials the proportionality between **B** and **H** holds up to very large induction values, and $\mu \gg 1$. That is why a core made of such a material increases the self-inductance correspondingly.

Forces Acting on a Conductor in a Magnetic Field. The expression for the force acting upon a volume element of a medium in a magnetic field is in the most general case fairly complex. However, if the relative magnetic permeability is close to unity, the force, in complete analogy with the magnetic component of the Lorentz force [14.29], is

$$d\mathbf{F} = \frac{dV}{c}\,\mathbf{j}\times\mathbf{H} \tag{34.39}$$

Of greatest interest is the case of a linear conductor. Carrying out the substitution (34.13), we find the resultant force acting on a circuit as a whole:

$$\mathbf{F} = \frac{I}{c}\int d\mathbf{l}\times\mathbf{H} \tag{34.40a}$$

This integral takes into account a magnetic field of dual origin: applied to the circuit from outside and produced by the current I itself. But the latter cannot produce a resultant force acting on the circuit, since otherwise it would be able to make itself move in space as a whole, which would contradict the momentum conservation law. Separate parts of a circuit can, of course, cause one another to move.

Denoting the external field $\mathbf{H}_0$, we write (34.40a) as

$$\mathbf{F} = \frac{I}{c}\int d\mathbf{l}\times\mathbf{H}_0 \tag{34.40b}$$

Let us apply the generalised Stokes' theorem to the integral (34.40b). An element $d\mathbf{l}$ of a closed circuit can be replaced by $d\mathbf{S}\times\nabla$, where $d\mathbf{S}$ is an element of the surface stretched over the circuit. In the most general case ∇ refers to the whole integrand, but here, obviously, only to $\mathbf{H}_0$.

Going over to the surface integral and expanding the double vector product, we write

$$\begin{aligned}\mathbf{F} &= \frac{I}{c}\int [-d\mathbf{S}\,\operatorname{div}\mathbf{H}_0 + \operatorname{grad}(d\mathbf{S}\cdot\mathbf{H}_0)]\\ &= \frac{I}{c}\int [-d\mathbf{S}\,\operatorname{div}\mathbf{H}_0 + (d\mathbf{S}\cdot\nabla)\,\mathbf{H}_0 + d\mathbf{S}\times\operatorname{curl}\mathbf{H}_0]\end{aligned}$$

But $\operatorname{div}\mathbf{H}_0 = 0$ and $\operatorname{curl}\mathbf{H}_0 = 0$. Therefore only the second term,

$$\mathbf{F} = \frac{I}{c}\int (d\mathbf{S}\cdot\nabla)\,\mathbf{H}_0 \tag{34.41}$$

remains.

If the external field changes only slightly within the boundaries of the circuit, $\mathbf{H}_0$ can in the first approximation be taken outside the integral sign. Then, recalling the definition of the magnetic

moment of a current (34.22), we obtain the approximate expression for the force acting on the circuit:

$$\mathbf{F} = (\mathbf{m} \cdot \nabla)\, \mathbf{H}_0 \tag{34.42}$$

The same formula occurs in the magnetostatics of point charges (see [17.35]).

The moment of the forces applied to a circuit by an external uniform magnetic field is

$$\mathbf{K} = \mathbf{m} \times \mathbf{H}_0 \tag{34.43}$$

This follows from the definition of the moment of a force, $\mathbf{K} = (I/c) \int \mathbf{r} \times (d\mathbf{l} \times \mathbf{H}_0)$, and the expression for the magnetic moment (34.20). Taking advantage of the fact that $\int d\mathbf{l} \cdot \mathbf{r} = 0$ and substituting $(1/2) \int [d\mathbf{l}\,(\mathbf{r} \cdot \mathbf{H}_0) - \mathbf{r}\,(d\mathbf{l} \cdot \mathbf{H}_0)]$ for $\int d\mathbf{l}\,(\mathbf{r} \cdot \mathbf{H}_0)$, we arrive at Eq. (34.43), which is fundamental for an understanding of the principle of operation of electric motors.

EXERCISES

1. A magnetic field is produced by a system of parallel currents of infinite length parallel to the z axis. The magnetic permeability of the medium is constant. Write the basic magnetic field equations and establish the similarity with electrostatic field.

Solution. We select a vector potential A directed along the z axis and dependent only upon x and y. Condition (34.6) is thus satisfied automatically. The induction components are

$$B_x = \frac{\partial A}{\partial y}, \qquad B_y = -\frac{\partial A}{\partial x}$$

Equation (34.1) is then written as follows:

$$\frac{\partial^2 A}{\partial x^2} + \frac{\partial^2 A}{\partial y^2} = -\frac{4\pi j \mu}{c}$$

This scalar equation for a plane problem is similar to the electrostatic equation $\nabla^2 \varphi = -4\pi\rho_{ext}/\varepsilon$, where ρ_{ext} is the density of external charges introduced into the dielectric. For a two-dimensional problem we must choose

$$D_x = \varepsilon E_x = -\varepsilon \frac{\partial \varphi}{\partial x}, \qquad D_y = \varepsilon E_y = -\varepsilon \frac{\partial \varphi}{\partial y}$$

Consequently, μ/c in the Laplace equation is replaced by $1/\varepsilon$.

2. The current density has only an azimuthal component in a cylindrical coordinate system and is dependent on r and z: $j_r = j_z = 0$, $j_\varphi = j(r, z)$. Write the equation for the vector potential.

Solution. We assume that only the component $A_\varphi = A(r, z)$ of the vector potential is not zero. The expressions for the induction components are:

$$B_r = -\frac{\partial A}{\partial z}, \qquad B_z = \frac{1}{r}\frac{\partial}{\partial r}(rA)$$

After a second calculation of the curl in cylindrical coordinates, we arrive at the equation

$$\frac{\partial}{\partial r}\left(\frac{1}{r}\frac{\partial}{\partial r} rA\right) + \frac{\partial^2 A}{\partial z^2} = -\frac{4\pi\mu}{c} j(r, z)$$

3. Show that the magnetic induction of a linear conductor in a medium with a constant value of μ can be expressed in terms of the gradient of a scalar that is multiple valued when passing through a circuit linking the conductor.

Solution. The vector potential (34.14) of the circuit transforms into a surface integral according to the generalized form of Stokes' theorem:

$$\mathbf{A} = \frac{\mu}{c} I \int \frac{d\mathbf{l}}{|\mathbf{r}-\mathbf{r}'|} = \frac{\mu I}{c} \int d\mathbf{S} \times \operatorname{grad} \frac{1}{|\mathbf{r}-\mathbf{r}'|}$$

The surface must be drawn so that vector $\mathbf{r}$ does not lie in it. Hence the induction is equal to

$$\mathbf{B} = \operatorname{curl} \mathbf{A} = \frac{\mu I}{c} \int \operatorname{curl}\left(d\mathbf{S} \times \operatorname{grad} \frac{1}{|\mathbf{r}-\mathbf{r}'|}\right)$$

$$= \frac{\mu I}{c} \int \left[d\mathbf{S}\nabla^2 \frac{1}{|\mathbf{r}-\mathbf{r}'|} - (d\mathbf{S}\cdot\nabla) \operatorname{grad} \frac{1}{|\mathbf{r}-\mathbf{r}'|}\right]$$

The first term under the integral sign is equal to zero, since point $\mathbf{r}$ was selected not on the surface. The gradient operator applied to $|\mathbf{r} - \mathbf{r}'|^{-1}$ in the second term is taken outside the integral sign, because here integration is carried out with respect to $\mathbf{r}'$, while the gradient refers to point $\mathbf{r}$. Consequently

$$\mathbf{B} = -\operatorname{grad} \frac{\mu I}{c} \int (d\mathbf{S}\cdot\nabla) \frac{1}{|\mathbf{r}-\mathbf{r}'|}$$

Let us now determine the meaning of the scalar quantity:

$$-\int (d\mathbf{S}\cdot\nabla) \frac{1}{|\mathbf{r}-\mathbf{r}'|} = \int \frac{d\mathbf{S}(\mathbf{r}-\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|^3} = \int \frac{dS \cos\alpha}{(\mathbf{r}-\mathbf{r}')^2}$$

Here α is the angle between the normal to the surface element $d\mathbf{S}$ and the line drawn from $d\mathbf{S}$ to point $\mathbf{r}$. Therefore, $dS \cos\alpha$ is the projection of the surface element $d\mathbf{S}$ on a plane normal to the line, and the integrand as a whole gives the solid angle $d\Omega$ at which $d\mathbf{S}$ is seen from point $\mathbf{r}$ [12.27]. In passing along the closed circuit linked with the conductor, we see that

the solid angle changes by 4π on return to the initial point. Thus

$$\mathbf{B} = \text{grad}\, \frac{\mu I}{c}\, \Omega$$

and the potential $\mu I \Omega / c$ is a multiple-valued function.

4. Show that the resultant force and the moment of force with which a curcuit's own magnetic field acts on that circuit in a homogeneous medium are zero.

Solution. The resultant force is

$$\mathbf{F} = \frac{1}{c} \int (\mathbf{j} \times \mathbf{H})\, dV = -\frac{1}{4\pi} \int (\mathbf{H} \times \text{curl}\, \mathbf{H})\, dV$$

We make use of the equation

$$\text{grad}\left(\frac{\mathbf{H}^2}{2}\right) = (\mathbf{H}\cdot\nabla)\, \mathbf{H} \times \mathbf{H} \times \text{curl}\, \mathbf{H}$$

The integral of grad $\mathbf{H}^2/2$ over the volume transforms directly into a surface integral, and on a sufficiently distant surface becomes zero. The integral of $(\mathbf{H}\cdot\nabla)\, \mathbf{H}$ transforms by parts as follows:

$$\int (\mathbf{H}\cdot\nabla)\, \mathbf{H}\, dV = \int (\mathbf{H}\cdot d\mathbf{S})\, \mathbf{H} - \int \mathbf{H}\, \text{div}\, \mathbf{H}\, dV$$

The surface integral again becomes zero, and in a homogeneous medium div $\mathbf{B} = \mu$ div $\mathbf{H} = 0$.

The moment of force is

$$\mathbf{K} = \int \mathbf{r} \times (\mathbf{H} \times \text{curl}\, \mathbf{H})\, dV = \int \mathbf{r} \times \left[\frac{1}{2}\, \text{grad}\mathbf{H}^2 - (\mathbf{H}\cdot\nabla)\, \mathbf{H}\right] dV$$

Further, taking advantage of the fact that curl $\mathbf{r} = 0$ and integrating by parts, we obtain $\int (\mathbf{r} \times \text{grad}\, \mathbf{H}^2)\, dV = 0$. Similarly, again taking advantage of the fact that curl $\mathbf{r} = 0$, we find that the integral of the second term is equal to zero.

5. Show that the forces with which the magnetic field of a current-carrying circuit with a constant current acts upon that circuit tend to stretch the area of the circuit.

Solution. The independent variable in the expression for the differential of work (34.25) is the vector potential **A**. The same is true of (34.26). Since the current must be kept constant, it should be chosen as the independent variable. For that subtract from F_m the integral $c^{-1} \int (\mathbf{A}\cdot\mathbf{j})\, dV$, which yields $-F_m$. Hence, for one conductor

$$-F_m = -\frac{1}{2} \mathcal{L} I^2$$

This expression must have a minimum in steady current conditions, so that equilibrium corresponds to the greatest self-inductance. But since $\Phi = \mathcal{L} I$, the maximum value of $\mathcal{L}$ also yields the maximum magnetic flux that permeates the circuit in the case of steady current. Consequently, the magnetic

forces stretch the circuit. Note that there is a complete analogy between linear currents and vortices in an ideal fluid (Sec. 15). A vortex ring also expands, as can be seen from the example of smoke rings.

6. Two identical circular rings of radius r lie in parallel planes at a distance $2a$ apart. Determine the coefficient of mutual inductance between them in the form of a definite integral.

Answer.

$$\mathscr{L}_{12}=\frac{2\pi\mu r^2}{c^2}\int\limits_0^{\pi/2}\frac{d\alpha}{(a^2+r^2\sin^2\alpha)^{1/2}}$$

35

QUASI-STATIONARY CURRENTS

Quasi-stationary Conditions. Up till now, in considering a field in a medium, we always assumed it to be constant in time. In such conditions either statistical equilibrium sets in (that is, definite values of magnetic and electric polarization as functions of the field) or the rate of change of electric polarization, characterized by the conduction current $\mathbf{j}$, becomes constant. When a field is switched on, equilibrium does not set in instantaneously but over a certain characteristic time, called the *relaxation time* of the system. The same is true of direct current. If the field changes insignificantly in that time, we can use the equilibrium quantities μ and ε and the electrical conductivity σ for direct current. In other words, we shall assume that all relationships between electromagnetic quantities of the type $\mathbf{B} = \mu\mathbf{H}$, $\mathbf{D} = \varepsilon\mathbf{E}$, and $\mathbf{j} + \sigma\mathbf{E}$ involve the same constants as in constant fields.

Let us write Maxwell's equations for such a slowly variable or *quasi-stationary* field, as it is conventionally called. The divergence equations remain the same and have the form (28.24) and (28.25). Equation (28.27), which involves time, also remains the same. In it $\mu\mathbf{H}$ must be substituted for $\mathbf{B}$ with the static value of magnetic permeability. Introducing the conduction current according to (31.3) into (28.28), we obtain

$$\operatorname{curl}\mathbf{H}=\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}+\frac{4\pi\mathbf{j}}{c}=\frac{1}{c}\frac{\partial\mathbf{E}}{\partial t}+\frac{4\pi\sigma\mathbf{E}}{c} \tag{35.1}$$

It was pointed out in Section 31 that for good conductors the value of σ is of the order 10^{18} s^{-1}. This means that the electric field would have to change in 10^{-18} seconds for the first term on the right

to become comparable with the second term, which does not involve time.

Defining a field by its frequency ω, we see that the displacement current is small if the inequality

$$\omega \ll \sigma \tag{35.2}$$

is satisfied.

Then the state of the system is defined by the instantaneous value of the field and not its time rate of change, which is involved in the quasi-stationary condition.

Furthermore, if the stationary value of conductivity is established in a system in time τ, then to be able to substitute the quantity σ for a constant field the inequality

$$\omega \ll \frac{1}{\tau} \tag{35.3}$$

must also be satisfied.

In the case of good conductors the satisfaction of the inequality (35.3) guarantees satisfaction of (35.2). In poor conductors it may be necessary, at some value of the field frequency, to take into account both terms on the right in (35.1). A variable field that is quasi-stationary in a metal may not be quasi-stationary in a semiconductor.

The definition of a quasi-stationary field also includes the requirement that it be in the same phase throughout the system. This in turn imposes limitations on the dimensions of the system: they must be small in comparison with the quantity $\lambda/(2\pi) = c/\omega$, where λ is the wavelength of the corresponding electromagnetic waves in vacuum. In a nonconducting medium the expression changes somewhat (see Sec. 36), but retains the same meaning. For a quasi-stationary field we have one more condition:

$$l \ll \frac{\lambda}{2\pi} = \frac{c}{\omega} \tag{35.4}$$

When a field is investigated on the boundary of a conductor, an additional requirement arises, which will be examined later in this section.

The Basic Equations of Quasi-stationary Fields. Thus, to satisfy the conditions listed above we have a set of equations

$$\operatorname{curl} \mathbf{H} = \frac{4\pi}{c} \mathbf{j} = \frac{4\pi\sigma}{c} \mathbf{E} \tag{35.5}$$

$$\operatorname{div} \mathbf{D} = 0 \tag{35.6}$$

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} \tag{35.7}$$

$$\operatorname{div} \mathbf{B} = 0 \tag{35.8}$$

Applying the operator curl to Eq. (35.5) and making use of (35.7), we obtain

$$\operatorname{curl}\operatorname{curl}\mathbf{H}=\frac{4\pi\sigma}{c}\operatorname{curl}\mathbf{E}=-\frac{4\pi\sigma}{c^2}\frac{\partial\mathbf{B}}{\partial t} \tag{35.9}$$

If the proportional dependence $\mathbf{B}=\mu\mathbf{H}$ holds, that is, if the medium is nonferromagnetic or, what is more important in practice, the system includes soft (easily magnetized) iron, then

$$\operatorname{curl}\operatorname{curl}\mathbf{H}=\operatorname{grad}\operatorname{div}\mathbf{H}-\nabla^2\mathbf{H}=-\frac{4\pi\sigma\mu}{c^2}\frac{\partial\mathbf{H}}{\partial t} \tag{35.10}$$

But if $\mathbf{H}=\mathbf{B}/\mu$, then $\operatorname{div}\mathbf{H}=0$, whence we obtain the magnetic field equation

$$\nabla^2\mathbf{H}=\frac{4\pi\sigma\mu}{c^2}\frac{\partial\mathbf{H}}{\partial t} \tag{35.11}$$

This equation is of the type of the diffusion equation (or the heat conductivity equation): see Exercise 3, Section 17. It permits certain conclusions based solely on considerations of dimension.

Let a conductive body of thickness l (its length is assumed greater) be placed in a region of space where a magnetic field is suddenly switched on or off. It is then possible to assess the time it takes for the field to penetrate the conductor or damp out in it. From Eq. (35.11) we see that the quantity

$$t_0=\frac{\sigma\mu l^2}{c^2} \tag{35.12a}$$

has the dimension of time.

The greater the conductivity the slower an external field penetrates a conductor. This is explained by Lenz's induction law: when the field is switched on, a current is induced in the conductor, which produces a reverse field. At high conductivity the current attenuates slowly. In a superconductor it does not attenuate at all. That is why a magnetic field does not penetrate a superconductor: a current appears in a thin surface layer which fully screens the field.

Let us now assume that a magnetic field on the surface of a conductor varies sinusoidally with a frequency ω. The currents induced in the conductor are directed in such a way so as to counteract the penetration of the field. As a result the magnetic field within the conductor will be other than zero only to a certain depth δ. This quantity is also easily evaluated from considerations of dimensions according to Eq. (35.11):

$$\delta\approx\left(\frac{c^2}{\sigma\mu\omega}\right)^{1/2} \tag{35.12b}$$

At very high frequencies or very great conductivities δ is small. A situation may develop in which δ becomes smaller than the free

path of an electron in the metal. But for such small lengths the very concept of conductivity loses meaning, just as the concept of viscosity is inapplicable to distances that are small in comparison with the free path of a gas molecule. From this we obtain one more criterion justifying the theory of quasi-stationary fields: the depth to which a field penetrates a conductor must be great in comparison with the free path of an electron in the metal. Otherwise the relationship $\mathbf{j} = \sigma\mathbf{E}$ cannot be used. Instead of differential equations we have to use an integral equation taking into account the balance of molecules coming to the surface of the metal and scattered back into it.

Foucault Currents. We shall consider two limiting cases: when the dimensions of the conductor are small in comparison with the penetration depth δ, and when they are great. In the first case the currents induced by a variable field screen it weakly. Their own magnetic field can therefore be neglected, and the right-hand side of Eq. (35.7) is assumed to involve only the magnetic induction $\mathbf{B}_0$ due to the magnetic field applied from outside:

$$\operatorname{curl} \mathbf{E} = -\frac{1}{c}\frac{\partial \mathbf{B}_0}{\partial t} \tag{35.13}$$

If the magnetic permeability of a conductor is great, $\mathbf{B}_0$ is determined from the magnetostatic case; in the case of a nonmagnetic substance $\mathbf{B}_0$ is replaced by the external field $\mathbf{H}_0$. Together with the equation $\operatorname{div} \mathbf{E} = 0$, which follows from (35.6), it fully defines the induced electric field. Thereby the currents induced by the external magnetic field are also known. They are called *Foucault*, or *eddy*, *currents*. In bulk conductors with large cross-sectional areas these may be considerable indeed. That is why transformer cores, dynamo armature, etc., are made of iron lamina separated by insulating layers.

Let us determine how the energy dissipation of an external magnetic field depends on its oscillation frequency. It follows from Eq. (35.13) that an induced electric field is proportional to the frequency, because the first time derivative appears in the right-hand side. The current density $\mathbf{j} = \sigma\mathbf{E}$ is thus also proportional to the frequency, while Joule losses $\mathbf{jE}$ depend upon the frequency according to a quadratic law.

The Skin Effect. In the reverse case, when the depth of penetration of a variable field into a conductor is small, we shall for the sake of simplicity assume the conductor to be infinitely large. Let the conductor surface lie in the xy-plane and the z axis be directed inside. Then the current has only one component along the x axis, and the magnetic field has one along the y axis. We assume

that the field varies with time according to a harmonic law, which is conveniently taken in complex form:

$$H_y = He^{-i\omega t}, \qquad H = H(z) = H_0 e^{z/\delta}$$

Substituting H_y into (35.11), we obtain

$$\frac{d^2H}{dz^2} = -\frac{4\pi\sigma\mu\omega i}{c^2} H \equiv -i\lambda^2 H$$

$$\delta^{-1} = \pm(-i)^{1/2}\lambda \tag{35.14}$$

Of the two solutions of this equation we should choose the one that falls off inside the conductor. Representing the factor $-i$ as $e^{-i\pi/2}$, we have

$$-(-i)^{1/2} = e^{i\pi - i\pi/4} = e^{3\pi i/4} = \frac{-1+i}{\sqrt{2}}$$

The quantity δ is given by the formula

$$\delta = \frac{c}{(2\pi\sigma\mu\omega)^{1/2}} \tag{35.15}$$

Therefore the expression for the magnetic field in complex form has the form

$$H_y = H_0 e^{-i\omega t + iz/\delta} e^{-z/\delta} \tag{35.16}$$

On the surface of the metal (at $z = 0$) the field coincides with the external field. Hence, with the help of (35.5), we find the electric field

$$E_x = \frac{c}{4\pi\sigma} \operatorname{curl}_x \mathbf{H} = -\frac{c}{4\pi\sigma}\frac{dH_y}{dz}$$

The derivative of the magnetic field with respect to z introduces the phase factor $e^{3\pi i/4}$ into this expression. Taking into account the sign before the derivative, we find that this factor acquires the form $e^{i\pi + 3i\pi/4} = e^{7i\pi/4}$. Subtracting $2\pi i$ from the exponent (which can always be done), we obtain $e^{-i\pi/4}$. Hence the electric field is

$$E_x = \frac{cH_0}{4\pi\sigma\delta} e^{-i\pi/4} e^{-z/\delta} e^{-i\omega t - iz/\delta} \tag{35.17}$$

Taking now the real parts of (35.16) and (35.17), we finally obtain

$$H_y = H_0 e^{-z/\delta} \cos\left(\frac{z}{\delta} - \omega t\right) \tag{35.18}$$

$$E_x = \frac{cH_0}{4\pi\sigma\delta} e^{-z/\delta} \cos\left(\frac{z}{\delta} - \omega t - \frac{\pi}{4}\right) \tag{35.19}$$

Thus, the electric field and current $\mathbf{j} = \sigma\mathbf{E}$ have a phase lag of $\pi/4$ with respect to the magnetic field.

Here the field and the current are concentrated in the external layer of the conductor, whence the name, *skin effect*.

The effect holds for conductors of circular cross section. If the penetration depth is appreciably less than the radius of a cylindrical conductor, the current does not flow across the whole section. That is why in such cases conductors are made either hollow or multiple-strand.

The energy of the electromagnetic field flows across the conductor's surface, evolving inside the conductor as Joule heat. The energy flux across a unit area is, as is known from [15.26], the normal component of the Poynting vector:

$$\mathbf{P} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H}$$

or, in the present case,

$$P = \frac{c}{4\pi} E_y H_x \qquad (35.20)$$

Substituting the real parts of (35.18) and (35.19), we obtain the expression for the instantaneous value of P as a function of time. Of greater interest is its value averaged over time. It is conveniently calculated with the help of the complex expressions (35.16) and (35.17). Their real parts have the form

$$H_y = \frac{1}{2} (\mathrm{h}e^{-i\omega t} + \mathrm{h}^* e^{i\omega t}), \qquad E_x = \frac{1}{2} (\mathrm{e}e^{-i\omega t} + \mathrm{e}^* e^{i\omega t})$$

In developing the mean of the product $H_y E_x$ we must retain only the terms not dependent on time:

$$\overline{H_y E_x} = \frac{1}{4} (\mathrm{he}^* + \mathrm{h}^*\mathrm{e}) = \frac{1}{2} \operatorname{Re} (\mathrm{he}^*) \qquad (35.21)$$

This quantity must be taken at $z = 0$, that is, on the surface of the conductor. Hence

$$P = \frac{c}{4\pi} \frac{c}{4\pi\sigma\delta} \frac{1}{2\sqrt{2}} H_0^2 \qquad (35.22)$$

Comparing this expression with (35.15), we conclude that the losses are proportional to the square root of the frequency of the field.

Resistance in an A-C Circuit. Let us consider a circuit comprising a resistance, a capacitance, and an inductance joined in series. We assume capacitance C and inductance L to be concentrated in a capacitor and a coil, respectively. Let the external magnetic flux passing through the coil be varying according to a given law. An electromotive force develops in the circuit, which is expressed with

the help of (35.13) as follows:

$$\mathscr{E} \equiv \int \mathbf{E}\, d\mathbf{l} = \int \operatorname{curl} \mathbf{E}\, d\mathbf{S} = -\frac{1}{c}\frac{\partial}{\partial t}\int \mathbf{B}_0\, d\mathbf{S}$$
$$= -\frac{1}{c}\frac{\partial \Phi_0}{\partial t} \tag{35.23}$$

The work done by the electromotive force in unit time is equal to $I\mathscr{E}$. This work is distributed as follows. The part RI^2 irreversibly transforms into Joule heat, another part changes the magnetic energy $\mathscr{L}I^2/2$ of the coil, and another is expended on changing the electrostatic energy $e^2/(2C)$ of the capacitor (where e is the instantaneous value of the plate charges). Hence, the energy balance can be written as follows:

$$\mathscr{E}I = RI^2 + \frac{d}{dt}\frac{\mathscr{L}I^2}{2} + \frac{d}{dt}\frac{e^2}{2C} \tag{35.24}$$

The current flowing to the capacitor is connected with the charges on its plates by the relationship

$$I = \frac{de}{dt} \tag{35.25}$$

Performing the differentiation in the third term in the right-hand side of (35.24) and cancelling out I, we obtain the nonhomogeneous equation

$$\mathscr{L}\frac{dI}{dt} + RI + \frac{e}{C} = \mathscr{E} \tag{35.26}$$

where charge e is connected with current I by the relationship (35.25). Thus, the instantaneous value of I depends not only upon the electromotive force at the given time but on the whole history of its change as well.

In practice, however, the emf usually varies with time according to a harmonic law, that is, sinusoidally. Since Eq. (35.26) is linear, the current generated by that emf also obeys a harmonic law. The natural oscillations in the circuit, which satisfy the equation without the right-hand side, attenuate with time due to Ohmic resistance. Therefore, if $\mathscr{E} = \mathscr{E}_0 e^{-i\omega t}$, then a certain time after the emf was switched on the current in the circuit begins to alternate according to a similar law: $I = I_0 e^{-i\omega t}$. At the same time

$$\frac{dI}{dt} = -i\omega I, \qquad e = \frac{I}{-i\omega}$$

Hence, the differential equation (35.26) becomes an algebraic equation

$$\left(R - i\omega\mathscr{L} + \frac{i}{\omega C}\right) I = \mathscr{E} \tag{35.27}$$

The coefficient of I is the *complex impedance*, or simply the *impedance*:

$$Z(\omega) \equiv R - i\left(\omega\mathscr{L} - \frac{1}{\omega C}\right)$$
$$= \left[R^2 + \left(\omega\mathscr{L} - \frac{1}{\omega C}\right)^2\right]^{1/2} e^{i\psi} \tag{35.28}$$

where the phase lag, ψ, of the current relative to the emf is determined by the formula

$$\tan\psi = \frac{1}{R}\left(\frac{1}{\omega C} - \omega\mathscr{L}\right) \tag{35.29}$$

Inductance results in phase lag of the current, capacitance in phase lead.

A System of Coupled Circuits. Suppose now that a given circuit is inductively coupled with other circuits. According to (34.33), they induce in it a magnetic flux

$$\Phi_i' = C \sum_{k \neq i} \mathscr{L}_{ik} I_k \tag{35.30}$$

This sum does not include the natural magnetic flux of the given circuit. Then, according to the last equality in (35.23), the emf in the ith circuit is

$$\mathscr{E}_i = \mathscr{E}_{i(\text{s.i})} - \frac{1}{C}\frac{\partial \Phi_i'}{\partial t} = \mathscr{E}_{i(\text{s.i})} - \sum_{k \neq i} \mathscr{L}_{ik}\frac{dI_k}{dt} \tag{35.31}$$

We write the partial derivative of the magnetic flux Φ_i' because the flux can vary in a steady magnetic field owing to the motion of the circuit; the meaning here is that the variation in flux is due to variation of the field. Substituting the expression for emf into Eq. (35.26), we obtain a set of ordinary differential equations

$$\sum_k \mathscr{L}_{ik}\frac{dI_k}{dt} + R_i I_i + \frac{e_i}{C_i} = \mathscr{E}_{i(\text{s.i})} \tag{35.32}$$

where the sum is now taken over all k's, including i.

If all the $\mathscr{E}_{i(\text{s.i})}$ vary according to the same harmonic law, that is, proportional to $e^{-i\omega t}$, the set (35.32) becomes an algebraic system of equations.

This makes it possible to calculate linear sinusoidal a-c circuits along the same lines as d-c circuits. Ohm's law for each closed circuit is written in terms of the impedance matrix, Z_{ik}, defined as

$$Z_{ik}\omega = \delta_{ik}R_i - i\left(\omega\mathscr{L}_{ik} - \frac{\delta_{ik}}{\omega C_i}\right) \tag{35.33}$$

In calculations, Kirchhoff's first law, which states that the algebraic sum of the currents which meet at a junction point of an electric circuit is zero, is also employed.

Note that alternating current passes through capacitors, as the change of charge on one plate causes a corresponding change of charge on the other. Irreversible losses occur on resistors.[5]

The Mechanical Analogue of an A-C Circuit. Equations (35.32) can be treated as a set of the Lagrange equations of a mechanical model whose generalized coordinates are equal to the charges e_i, and the generalized velocities to the currents $I_i = \dot{e}_i$. The corresponding Lagrange function has the form

$$L = \frac{1}{2}\sum_{i,k} \mathscr{L}_{ik} I_i I_k - \frac{1}{2}\sum_i \frac{e_i^2}{C_i} + \sum_i e_i \mathscr{E}_{i(\mathrm{s.i})} \tag{35.34}$$

In addition, the so-called *dissipation function*

$$D = \frac{1}{2}\sum_i R_i I_i^2 \tag{35.35}$$

is determined in such a way that the Lagrange equations, when dissipation occurs, have the form

$$\frac{d}{dt}\frac{\partial L}{\partial I_i} - \frac{\partial L}{\partial e_i} = -\frac{\partial D}{\partial I_i} \tag{35.36}$$

which is equivalent to (35.32).

Of course, such equations cannot be derived from the variation principle [Sec. 2] alone. In the most general case the left-hand side of (35.36) describes purely mechanical properties of a system, whereas dissipation is essentially a statistical, irreversible process. As compared with the elementary mechanical law expressed with the help of the variation principle, involved here is an additional assumption in the form of the statistical law in the right-hand side of (35.36). The equations of mechanics can also be written in this way if friction forces proportional to the velocities are acting on the system.

The establishment of mechanical similarity of electric oscillations in circuits played an important part in the general development of physics. It was found that if quantitative description is involved, an electromagnetic field can be treated as a mechanical system. This is a manifestation of the unity of the laws of nature. Such an approach yielded the general equations of microscopic electrodynamics, that is, Maxwell's equations [Sec. 15].

---

[5] Actually, losses also occur in the dielectric materials of capacitors and in the ferromagnetic materials of induction coils, due to relaxation processes (see Sec. 36).

Motion of a Conductor in Magnetic Field. Let us determine the electromotive force in a circuit moving in a magnetic field. In a circuit at rest it is expressed by formula (35.23). But if some part of the circuit is moving in space, an additional electric field may appear in a reference frame connected with it if the magnetic induction **B** was not zero in the initial frame.

The electric field can be easily determined from the Lorentz transformation formulas for a field in a medium (Exercise 2, Section 28). Considering the velocity v of the conductor to be small in comparison with the speed of light, and remembering that it was assumed to be directed along the x axis, we rewrite the transformation formulas as follows:

$$E'_x = E_x, \qquad E'_y = E_y - \frac{v_x}{c} B_z$$

$$E'_z = E_z + \frac{v_x}{c} B_y \tag{35.37}$$

They can be brought together into one expression, written in vector form as follows:

$$\mathbf{E}' = \mathbf{E} + \frac{1}{c} \mathbf{v} \times \mathbf{B} \tag{35.38}$$

Such an electric field acts within a conductor having the velocity **v**. Therefore the emf in the circuit is

$$\mathscr{E} = \int (\mathbf{E}' \cdot d\mathbf{l}) = \int (\mathbf{E} \cdot d\mathbf{l}) + \int \frac{1}{c} (\mathbf{v} \times \mathbf{B})\, d\mathbf{l} \tag{35.39}$$

The first integral is expressed in terms of the change of the magnetic flux in the stationary circuit (35.23). In the second integral we perform a cyclic permutation:

$$(\mathbf{v} \times \mathbf{B}) d\mathbf{l} = (d\mathbf{l} \times \mathbf{v}) \mathbf{B}$$

But the vector product involved here is the area swept out by the line segment $d\mathbf{l}$ in unit time in the motion of the circuit, that is,

$$d\mathbf{l} \times \mathbf{v} = -\frac{d\mathbf{S}}{dt} \tag{35.40}$$

The origin of the minus sign in this equation is apparent from Figure 46. Vector $d\mathbf{l} \times \mathbf{v}$ is directed out of the page, whereas the vector of the area, according to the selected direction along the circuit, is into the page. Consequently, the sign of the area increment is opposite that of the area vector, which justifies the sign in the expression (35.40).

Transforming the first integral (35.39) according to Stokes' theorem, we obtain

$$\mathscr{E} = -\frac{1}{c} \int \frac{\partial \mathbf{B}}{\partial t} d\mathbf{S} - \frac{1}{c} \int \mathbf{B} \frac{d\mathbf{S}}{dt} = -\frac{1}{c} \frac{d\Phi}{dt} \tag{35.41}$$

Here the sign of the total derivative with respect to time emphasises that the total change is taken of the magnetic flux through the circuit due to both the change of vector **B** with time and the motion of the circuit in space.

It is apparent from Eq. (35.39) that the second term does not vanish only if **B** is not parallel to **v**. In other words, electromotive

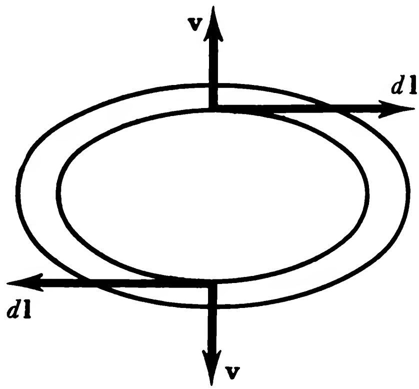


Figure 46

force appears because the circuit "crosses" the magnetic induction lines, although these lines represent a purely conventional construction in the sense as the meridians and parallels on the surface of the earth.

EXERCISES

1. Determine the complex impedance of two circuits joined in parallel.

Answer.

$$Z^{-1}=Z_1^{-1}+Z_2^{-1}$$

2. Determine the natural current oscillations in a circuit.

Solution. Assuming that $\mathscr{E}=0$ in (35.27), we find the equation of natural oscillations:

$$Z(\omega)=R-i\omega\mathscr{L}+\frac{i}{\omega C}=0$$

whence

$$\omega=-\frac{iR}{2\mathscr{L}}\pm\left(\frac{1}{\mathscr{L}C}-\frac{R^2}{4\mathscr{L}^2}\right)^{1/2}$$

At $R < 2(\mathscr{L}/C)^{1/2}$ the root is real, and the oscillations have a damping factor $R/(2L)$; otherwise the current attenuates aperiodically with two damping factors

$$\frac{R}{2\mathscr{L}} \pm \left(\frac{R^2}{4\mathscr{L}^2} - \frac{1}{\mathscr{L}C}\right)^{1/2}$$

3. Obtain the equation for the natural frequencies of a circuit made up of two circuits with impedances $Z_1(\omega)$ and $Z_2(\omega)$ joined in parallel.
Answer.

$$Z(\omega) = [Z_1^{-1}(\omega) + Z_2^{-1}(\omega)]^{-1} = 0$$

4. Two low-resistance circuits are inductively linked by a mutual inductance coefficient $\mathscr{L}_{12}$. Determine the natural frequencies and the approximate values of the damping factors of the oscillations.

Solution. Equations (35.32) are written as follows:

$$-i\omega\mathscr{L}_{11}I_1 - i\omega\mathscr{L}_{12}I_2 + R_1I_1 - \frac{I_1}{i\omega C_1} = 0$$

$$-i\omega\mathscr{L}_{12}I_1 - i\omega\mathscr{L}_{22}I_2 + R_2I_2 - \frac{I_2}{i\omega C_2} = 0$$

This set of equations has a unique solution only if the system determinant is equal to zero:

$$\begin{vmatrix} -i\omega\mathscr{L}_{11} + R_1 - \dfrac{1}{i\omega C_1} & -i\omega\mathscr{L}_{12} \\ -i\omega\mathscr{L}_{12} & -i\omega\mathscr{L}_{22} + R_2 - \dfrac{1}{i\omega C_2} \end{vmatrix} = 0$$

This is an equation of the fourth power in ω:

$$\omega^4(\mathscr{L}_{11}\mathscr{L}_{22} - \mathscr{L}_{12}^2) + i\omega^3(\mathscr{L}_{11}R_2 + \mathscr{L}_{22}R_1)$$
$$-\omega^2\left(\frac{\mathscr{L}_{11}}{C_2} + \frac{\mathscr{L}_{22}}{C_1} + R_1R_2\right) - i\omega\left(\frac{R_1}{C_2} + \frac{R_2}{C_1}\right) + \frac{1}{C_1C_2} = 0$$

In the zero approximation with respect to R_1 and R_2 we obtain a biquadratic equation with the following roots:

$$(\omega_\pm^2)_0 = \frac{\mathscr{L}_{11}C_1 + \mathscr{L}_{22}C_2}{2(\mathscr{L}_{11}\mathscr{L}_{22} - \mathscr{L}_{12}^2)C_1C_2}$$
$$\pm \frac{[(\mathscr{L}_{11}C_1 + \mathscr{L}_{22}C_2)^2 - 4(\mathscr{L}_{11}\mathscr{L}_{22} - \mathscr{L}_{12}^2)C_1C_2]^{1/2}}{2(\mathscr{L}_{11}\mathscr{L}_{22} - \mathscr{L}_{12}^2)C_1C_2}$$

If the mutual inductance coefficient $\mathscr{L}_{12}$ is equal to zero, the solution yields the frequency of independent oscillations. Of course this is also true of a precise equation, since the determinant is equal to the product of the diagonal elements. To find the damping factors in the first approximation we write

$$\omega_\pm^2 = (\omega_\pm^2)_0 + \Delta_\pm^2$$

Substituting this into the equation and retaining the terms linear in $\Delta_\pm^2$, R_1, and R_2, we obtain

$$\Delta_\pm^2 = -i\,(\omega_\pm)_0 \frac{(\omega_\pm^2)_0\, C_1C_2\,(\mathcal{L}_{11}R_2+\mathcal{L}_{22}R_1)-(R_1C_1+R_2C_2)}{2\,(\omega_\pm^2)_0\, C_1C_2\,(\mathcal{L}_{11}\mathcal{L}_{22}-\mathcal{L}_{22}^2)-(\mathcal{L}_{11}C_1+\mathcal{L}_{22}C_2)}$$

The denominator of this fraction is equal to

$$\pm\frac{1}{C_1C_2}\,[(\mathcal{L}_{11}C_1+\mathcal{L}_{22}C_2)^2-4\,(\mathcal{L}_{11}\mathcal{L}_{22}-\mathcal{L}_{12}^2)\,C_1C_2]^{1/2}$$

where the lower sign corresponds to $(\omega_-)_0$.

Let us show that Δ_-^2 involves the factor $-i$. Then we see that $-i\{(\omega_-)_0 - i\Delta_-^2/[2\,(\omega_-)_0]\}\,t = -i\,(\omega_-)_0t - \Delta_-^2t/[2\,(\omega_-)_0]$ occurs in the exponent, as it should.

The inequality that has to be proved has the form

$$(\omega_-^2)_0\,(\mathcal{L}_{11}R_1+\mathcal{L}_{22}R_2) < R_1C_1+R_2C_2$$

where $(\omega_-^2)_0$ was defined earlier. Since R_1 and R_2 are positive quantities, it is sufficient to assume one of them, say R_2, zero and perform the calculations for the other. After that, reducing the factors of R_1 to a common denominator and transposing terms, we arrive at the inequality, after substituting the explicit expression for $(\omega_-^2)_0$,

$$2\frac{\mathcal{L}_{12}^2}{\mathcal{L}_{22}^2} < \left[4\,\frac{\mathcal{L}_{12}^2}{\mathcal{L}_{22}^2}+\left(\frac{C_2}{C_1}-\frac{\mathcal{L}_{11}}{\mathcal{L}_{22}}\right)^2\right]^{1/2}-\left(\frac{C_2}{C_1}-\frac{\mathcal{L}_{11}}{\mathcal{L}_{12}}\right)$$

Getting rid of the irrationality, we see that the condition for satisfying the initial inequality reduces to

$$\mathcal{L}_{11}\mathcal{L}_{22}-\mathcal{L}_{12}^2>0$$

Satisfaction of this inequality is a necessary condition, since the quantity F_m, defined in (34.29), is essentially positive [7.18].

5. A conducting disk of radius a rotates in a uniform external magnetic field perpendicular to its plane, at an angular velocity ω. Determine the potential difference between the centre and the edge of the disk.

Solution. From (35.18), in a reference frame connected with the disk the electric field at distance r from the axis of rotation is

$$E=\frac{vH}{c}=\frac{\omega_0 rH}{c}$$

Hence, the potential difference is

$$\varphi=\int_0^a E\,dr=\frac{\omega_0a^2H}{2c}$$

6. A uniformly magnetized conducting sphere of radius a and moment $\mathbf{m}$ rotates at an angular velocity ω_0 about an axis parallel to the direction of the moment. Sliding contacts connected to a conductor are applied to its pole and equator. Determine the emf in a circuit comprising the conductor and the part of the sphere between the contacts (*unipolar induction*).

Solution. In a reference frame connected with the conductor the sphere is polarized not only magnetically but electrically as well:

$$\mathbf{P}=\frac{1}{c}\mathbf{v}\times\mathbf{m}$$

This can be seen from the general transformation equations (see Exercise 2, Section 28) in the limit $(1-v^2/c^2)^{-1/2}\equiv\alpha\to 1$. An electric field produced by a spatially distributed, time-constant electric moment **P** is electrostatic. Therefore the potential difference between the two fixed points (the contacts) does not depend on the path inside the conducting sphere. We can thus choose the path most suitable for integrating the electric field along the circuit, namely, a meridian from the pole to the equator. If **B** is the magnetic induction in the sphere, the electric field is $c^{-1}\mathbf{v}\times\mathbf{B}$. Hence only the normal component of the induction makes a contribution along the integration path, since **v** is everywhere directed tangent to the sphere and along the parallels.

The magnetic field on the surface of the sphere is

$$\mathbf{H}=\frac{3\mathbf{r}(\mathbf{m}\cdot\mathbf{r})-\mathbf{m}r^2}{r^5}$$

(see Exercise 2, Section 30).

Hence the normal component is

$$H_n=\frac{2m\cos\vartheta}{a^3}$$

It must be equated to the normal component of the induction within the sphere, $B\cos\vartheta$, whence

$$B=\frac{2m}{a^3}$$

Consequently

$$\mathscr{E}=\int_0^{\pi/2}\frac{2m\cos\vartheta}{a^3}\,a\omega_0\times a\sin\vartheta\,d\vartheta=\frac{m\omega_0}{ac}$$

Thanks to the sliding contacts, this potential difference induces direct current in the stationary conductor.

36

RAPIDLY VARIABLE FIELDS

General Equations. A rapidly variable field in a medium is one which varies appreciably in the time required for relaxation processes to take place in that medium. In the case of a dielectric medium

it is the time required for statistical equilibrium to set in, and in a conductor it is the time needed for direct current to be established after an electric field is switched on. These times differ greatly for different media and for different relaxation processes in the same medium.

If polarization in a medium does not at any moment attain its equilibrium value corresponding to the given field, the instantaneous polarization depends not only on the value of the field at the given instant but also on the whole history of the field. Furthermore, if the field is highly nonuniform in space, the polarization at a given point will also depend on the field in the surrounding space.

Thus, for the case of an arbitrary rapidly variable field the general equations (28.27) and (28.28) are largely meaningless, since, given the arbitrary nature of the time-variation of the field, there is no direct functional dependence of the electric displacement on the field. Nevertheless, for weak fields we can write a linear integral formula of the form

$$D(t) = E(t) + \int_0^\infty E(t-\tau) f(\tau)\, d\tau \tag{36.1}$$

which refers only to an isotropic medium. The function $f(\tau)$ gives the electric field's contribution to the electric displacement at the given time t if at a preceding time τ the field was $E(t-\tau)$. It is clear from physical considerations that the quantity $f(\infty)$ is zero or, at the very least, finite (see Exercise 1), since it describes the contribution to the electric displacement made by a field applied infinitely long ago.

Thus, given an arbitrary field-time dependence, Maxwell's equations in a medium are integral equations. But there is also one very important case when the explicit time dependence of the field is excluded from Maxwell's equations altogether. Namely, let

$$E = E_0 e^{-i\omega t} \tag{36.2}$$

that is, the field varies with time according to a complex harmonic law. It is called a *monochromatic* field. Substituting (36.2) into (36.1), we obtain

$$D_0 e^{-i\omega t} = E_0 e^{-i\omega t} + E_0 e^{-i\omega t} \int_0^\infty e^{i\omega\tau} f(\tau)\, d\tau$$

We shall not separate the exponential in future. Then

$$D = \left(1 + \int_0^\infty e^{i\omega\tau} f(\tau)\, d\tau\right) E \tag{36.3}$$

Formally this dependence has the same form as the common displacement-field relationship (30.14):

$$D = \varepsilon(\omega)\, E \tag{36.4}$$

where

$$\varepsilon(\omega) = 1 + \int_0^\infty e^{i\omega\tau} f(\tau)\, d\tau \tag{36.5}$$

A similar equation can be written for magnetic permeability, though in most cases in rapidly variable fields $\mu = 1$. For the sake of generality, however, we shall not, for the time being, assume $\mu = 1$. The dependence of ε and μ upon frequency is called *dispersion*.

The partial time derivatives of quantities of the form (36.2) are simply replaced by the factor $-i\omega$:

$$\frac{\partial \mathbf{D}}{\partial t} = -i\omega \mathbf{D}, \qquad \frac{\partial \mathbf{B}}{\partial t} = -i\omega \mathbf{B} \tag{36.6}$$

Therefore, Maxwell's equations for monochromatic fields have the following form:

$$\operatorname{curl} \mathbf{H} = -\frac{i\omega}{c} \mathbf{D} = -\frac{i\omega\varepsilon(\omega)}{c} E \tag{36.7}$$

$$\operatorname{curl} \mathbf{E} = \frac{i\omega}{c} \mathbf{B} = \frac{i\omega\mu(\omega)}{c} H \tag{36.8}$$

$$\operatorname{div} \mu(\omega)\mathbf{H} = 0 \tag{36.9}$$

$$\operatorname{div} \varepsilon(\omega)\mathbf{E} = 0 \tag{36.10}$$

Note that in earlier books the substitution (36.6) was frequently not carried out explicitly. Instead, certain mean values of ε and μ for the selected frequency range were substituted into Maxwell's equations. This requires specific assumptions concerning the weak form of the dependence $\varepsilon(\mu)$ and $\mu(\omega)$, which is not always satisfied.

The Imaginary Part of ε and μ. As can be seen from definition (36.5), the dielectric constant in a rapidly variable field is a complex quantity. Its real part is

$$\operatorname{Re} \varepsilon = 1 + \int_0^\infty f(\tau) \cos \omega\tau\, d\tau \tag{36.11}$$

and its imaginary part is

$$\operatorname{Im} \varepsilon = \int_0^\infty f(\tau) \sin \omega\tau\, d\tau \tag{36.12}$$

where the function $f(\tau)$ must by definition (36.1) be real. This expression involves, of course, only real values of the field. It follows from definitions (36.11) and (36.12) that

$$\varepsilon = \varepsilon' + i\varepsilon'' \tag{36.13}$$

The imaginary part ε'' corresponds to a phase lag of the electric displacement from the field by the quantity arc tan $(\varepsilon''/\varepsilon')$.

The functions $\varepsilon'(\omega)$ and $\varepsilon''(\omega)$ satisfy all known parity relationships. Thus, ε' involves only $\cos \omega t$, and therefore

$$\varepsilon'(-\omega) = \varepsilon'(\omega) \tag{36.14}$$

The quantity ε'', for its part, involves $\sin \omega t$. Hence

$$\varepsilon''(-\omega) = -\varepsilon''(\omega) \tag{36.15}$$

Thus, as a whole

$$\begin{aligned} \varepsilon(-\omega) &= \varepsilon'(-\omega) + i\varepsilon''(-\omega) \\ &= \varepsilon'(\omega) - i\varepsilon''(\omega) = \varepsilon^*(\omega) \end{aligned} \tag{36.16}$$

Although negative frequency has no direct physical meaning, the relationships (36.14)-(36.16) are very important. Let us now examine the physical meaning of the imaginary parts ε'' and μ''. As is known from [15.26], the Poynting vector

$$\mathbf{P} = \frac{c}{4\pi} \mathbf{E} \times \mathbf{H} \tag{36.17}$$

defines the density of the energy flux of an electromagnetic field. Its divergence is, obviously, equal to the density of the energy evolved per unit time for the given point. This conclusion can be drawn, for example, by analogy with the equation $\operatorname{div} \mathbf{j} = -\partial\rho/\partial t$ [12.18]. Averaging div **P** over time yields the density of the energy irreversibly evolved in the form of heat.

In order to take the mean value of a quantity quadratic with respect to the field we use a formula of the type (35.21):

$$\begin{aligned} \overline{\operatorname{div} \mathbf{P}} &= \frac{c}{4\pi} \frac{1}{2} \operatorname{Re} [\operatorname{div}(\mathbf{E} \times \mathbf{H}^*)] \\ &= \frac{c}{8\pi} \operatorname{Re} [\mathbf{H}^* \operatorname{curl} \mathbf{E} - \mathbf{E} \operatorname{curl} \mathbf{H}^*] \end{aligned}$$

Substituting curl **E** and curl **H*** from Maxwell's equations (36.7) and (36.8), we obtain

$$\overline{\operatorname{div} \mathbf{P}} = \frac{c}{8\pi} \operatorname{Re} [i\omega \mathbf{H}^* \mathbf{B} - i\omega \mathbf{E} \mathbf{D}^*] \tag{36.18}$$

Expressing ε and μ in the form (36.13), we finally obtain

$$\overline{\operatorname{div} \mathbf{P}} = \frac{c}{8\pi} \operatorname{Re} [i\omega \mathbf{H}^* (\mu' + i\mu'') \mathbf{H} - i\omega \mathbf{E} (\varepsilon' - i\varepsilon'') \mathbf{E}^*]$$
$$= -\frac{c\omega}{8\pi} (\mu'' |\mathbf{H}|^2 + \varepsilon'' |\mathbf{E}|^2) \qquad (36.19)$$

Since this formula describes irreversible losses, the imaginary parts of ε and μ, that is, ε'' and μ'', must be essentially positive functions of frequency at $\omega > 0$.

The Dielectric Constant at High and Low Frequencies. At very high oscillation frequencies the dielectric constant is easily calculated in general form. Indeed, in a rapidly variable electric field, in the time $2\pi/\omega$ no coupling forces acting on the electrons in matter will have time to affect their velocities appreciably. Consequently, the motion of an electron in the field of a wave is described by a simple equation, which involves only the forces produced by the external field:

$$m\ddot{\mathbf{r}} = e\mathbf{E}_0 e^{-i\omega t} \qquad (36.20)$$

It follows from this that

$$\mathbf{r} = -\frac{e}{m\omega^2} \mathbf{E}_0 e^{-i\omega t}$$

If there are N electrons in unit volume, their displacement under the action of the wave produces an electric polarization $\mathbf{P}$ such that

$$\mathbf{P} = Ne\mathbf{r} = -\frac{Ne^2}{m\omega^2} \mathbf{E}_0 e^{-i\omega t} \qquad (36.21)$$

The electric displacement $\mathbf{D}$ is expressed in terms of the polarization thus:

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \left(1 - \frac{4\pi Ne^2}{m\omega^2}\right) \mathbf{E} \qquad (36.22)$$

Comparing this expression with (36.4), we find the asymptotic value of the dielectric constant at high frequencies:

$$\varepsilon(\omega) = 1 - \frac{4\pi Ne^2}{m\omega^2} \qquad (36.23)$$

Note that it is a real quantity which is less than 1.

At very high frequencies, for example, in the X-ray range, the difference between metals and dielectrics disappears.

At low frequencies the complex quantity $\varepsilon(\omega)$ is substantially different for dielectric substances and metals. In the former case $\varepsilon(\omega)$ tends to its electrostatic value $\varepsilon(0) = \varepsilon_0$. A series expansion of the real part ε' in powers of ω^2 involves only even powers; a series

expansion of the imaginary part starts with the term proportional to ω and continues in odd powers, as follows from (36.14) and (36.15).

In the case of metals, at the lowest frequencies Eq. (34.1) should be used:

$$\operatorname{curl} \mathbf{H} = \frac{4\pi}{c}\mathbf{j} = \frac{4\pi\sigma}{c}\mathbf{E}$$

Comparing this equation with (36.7), we see that

$$-i\omega\varepsilon = 4\pi\sigma$$

or

$$\varepsilon = i\frac{4\pi\sigma}{\omega} \tag{36.24}$$

Thus, the expansion begins with a purely imaginary term which tends to infinity like ω^{-1} as $\omega \to 0$. This term is, as it should be, odd with respect to ω. The next term of the expansion is a real, constant quantity. It cannot, however, be given the meaning of a static dielectric constant, since metals lack equilibrium in a uniform electric field.

The Correspondence Between the Imaginary and Real Parts of ε. These important relationships were first derived by H.A. Kramers (1927) and R. de L. Kronig (1926) independently. Let us first examine expression (36.3), assuming ω to be a complex quantity $\omega' + i\omega''$:

$$\varepsilon(\omega) = 1 + \int_0^\infty \exp(i\omega'\tau - \omega''\tau) f(\tau)\, d(\tau) \tag{36.25}$$

The integral has a finite value at any positive value of ω'', because as $\tau \to \infty$ the factor $\exp(-\omega''\tau)$ tends to zero. By the condition imposed above on the function $f(\tau)$, it does not affect the divergence of the integral (36.25). Since $\exp(-\omega''\tau)$ decreases faster than any power of τ, all the derivatives of $\varepsilon(\omega)$ with respect to ω are also finite at $\omega'' > 0$. In other words, ε is defined as a function of the complex variable ω at all points of the half-plane $\omega'' > 0$, that is, above the real axis. It was stated in Section 15 that the derivative of a function of a complex variable should not depend on the direction in which it is taken. If a function $R(z)$ is the primitive of $Q(z)$, that is, if $Q(z) = dR/dz$, then the integral $R = \int_{z_0}^{z_1} Q(z)\, dz$ does not depend on the path between points z_0 and z_1 in the complex ω-plane. To see this it is sufficient to check whether $Q(z)\, dz$ is a total differential.

Representing Q as $Q' + iQ''$ and dz as $dx + i\,dy$, we obtain

$$Q\,dz = (Q' + iQ'')\,(dx + i\,dy) = (Q' + iQ'')\,dx + (iQ' - Q'')\,dy$$

The condition that $Q\,dz$ is a total differential (the integral of which is independent of the path) has, as is known, the following form:

$$\frac{\partial}{\partial y}(Q' + iQ'') = \frac{\partial}{\partial x}(iQ' - Q'') \tag{36.26}$$

After separating the real and imaginary parts, this yields the Cauchy-Riemann equations mentioned in Section 15. They must be satisfied at all points through which the integration path passes. If the path is deformed and passes through a point where the Cauchy-Riemann equations are not valid, the values of the integral before and after the point may be different.

Derivatives exist in all points of the domain through which the path passes if the function $Q(z)$ does not become infinite in it and is single-valued. Let us explain the meaning of the latter condition. For this consider a multiple-valued function $f(z) = (z - z_0)^\alpha$. If $z - z_0$ is represented as $|z - z_0|e^{i\psi}$, then $f(z) = |z - z_0|^\alpha\, e^{i\psi\alpha}$. In passing point $z = z_0$, 2π is added to the argument ψ, while $f(z)$ receives the factor $e^{2\pi i\alpha}$, which is not equal to unity if α is not an integer. Then if $dz = \lim (z - z_0)$ as $z \to z_0$, the differential dz is not single valued, so that the derivative depends on the direction of dz. If the path crosses such a point, the integral receives an arbitrary factor.

But in the upper half-plane ($\omega'' > 0$) the function $\varepsilon(\omega)$ does not have such points. Indeed, by differentiating $(\omega - \omega_0)^\alpha$ a sufficient number of times we finally arrive at a negative exponent, and the function becomes infinite at $\omega = \omega_0$. But as we have just showed, all derivatives of ε with respect to ω are finite. Hence $\varepsilon(\omega)$ is single valued in the upper half-plane.

Let us now consider the function

$$\frac{\varepsilon(\omega) - 1}{\omega - \omega_1}$$

at $\operatorname{Im}\omega_1 = 0$. For dielectric substances this function, like $\varepsilon(\omega)$, is finite and single valued above the real axis, while on the axis itself ($\omega'' = 0$) it becomes infinite at $\omega = \omega_1$. Let us take the integral of this function along the closed path shown in Figure 47. This path passes through the upper half-plane along an infinitely large semicircle, then along the real axis from $-\infty$ to the point $\omega = \omega_1 - \rho$, then along an infinitely small semicircle of radius ρ around the point $\omega = \omega_1$ above it and, finally, again along the real axis from point $\omega = \omega_1 + \rho$ to $\omega' = \infty$.

From what has just been proved, the integral under consideration is equal to zero, since it is taken along a closed path inside which

the integrand is everywhere single valued and analytic. The integral along the upper semicircle is itself equal to zero, since on it, from (39.23), the function $(\varepsilon - 1)/(\omega - \omega_1)$ tends to zero as ω^{-3}, while the circumference increases as ω.

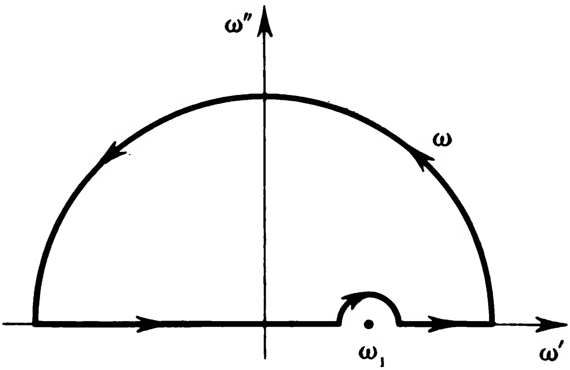


Figure 47

There remain the integrals along the real axis and small semicircle. The former is equal to

$$\lim_{\rho\to 0}\left(\int_{-\infty}^{\omega_1-\rho}\frac{\varepsilon-1}{\omega-\omega_1}\,d\omega+\int_{\omega_1+\rho}^{\infty}\frac{\varepsilon-1}{\omega-\omega_1}\,d\omega\right)$$

This integral tends to a finite limit as $\rho \to 0$, and is called the *principal value* of the integral. It is denoted by the letter P before the integral sign.

Let us first show that

$$\int_{-\infty}^{\infty}\frac{d\omega}{\omega-\omega_1}=\lim_{\substack{\Omega\to\infty\\ \rho\to 0}}\left\{\int_{-\Omega}^{\omega_1-\rho}\frac{d\omega}{\omega-\omega_1}+\int_{\omega_1+\rho}^{\Omega}\frac{d\omega}{\omega-\omega_1}\right\}=0$$

Indeed

$$\lim_{\substack{\Omega\to\infty\\ \rho\to 0}}\left\{-\int_{-\Omega}^{\omega_1-\rho}\frac{d\omega}{\omega_1-\omega}+\int_{\omega_1+\rho}^{\Omega}\frac{d\omega}{\omega-\omega_1}\right\}$$

$$=-\ln(\omega_1-\omega)\Big|_{-\Omega}^{\omega_1-\rho}+\ln(\omega-\omega_1)\Big|_{\omega_1+\rho}^{\Omega}$$

$$=\ln\left(\frac{\Omega-\omega_1}{\rho}\,\frac{\rho}{\Omega-\omega_1}\right)=\ln 1=0$$

The integral along a semicircle of radius ρ is taken as follows. Since on the semicircle

$$\omega-\omega_1=\rho e^{i\psi}, \qquad d\omega=i\rho\, d\psi e^{i\psi}, \qquad \varepsilon(\omega)=\varepsilon(\omega_1),$$

we find that

$$i\int_{\pi}^{0} d\psi\,[\varepsilon(\omega_1)-1]=-i\pi\,[\varepsilon(\omega_1)-1]$$

$$=-i\pi\,[\varepsilon'(\omega_1)+i\varepsilon''(\omega_1)-1]$$

The minus here is due to the fact that angle ψ is measured counterclockwise, while the semicircle is passed, as can be seen in Figure 47, clockwise.

Since the total integral is equal to zero, we obtain separate equations for the real and imaginary parts:

$$\pi\varepsilon''(\omega_1)=-\mathrm{P}\int_{-\infty}^{\infty}\frac{\varepsilon'(\omega)}{\omega-\omega_1}\,d\omega \tag{36.27}$$

$$\pi\,[\varepsilon'(\omega_1)-1]=\mathrm{P}\int_{-\infty}^{\infty}\frac{\varepsilon''(\omega)}{\omega-\omega_1}\,d\omega \tag{36.28}$$

These formulas are called *dispersion relations* or *Kramers-Kronig relations.*

The Meaning of the Dispersion Relations. Equation (36.1), which lies at the basis of the dispersion relations, expresses the causality principle: electric displacement is affected by the history of the field; it cannot be affected by the future values of the field. That is why under the integral sign stands the quantity $E\,(t-\tau)$, where $0\leqslant\tau\leqslant\infty$.

Let us now take advantage of the fact that $\varepsilon''(-\omega)=-\varepsilon''(\omega)$; see (36.15). Then formula (36.28) reduces to the form

$$\pi\,[\varepsilon'(\omega_1)-1]=\int_{-\infty}^{0}\frac{\varepsilon''(\omega)\,d\omega}{\omega-\omega_1}+\mathrm{P}\int_{0}^{\infty}\frac{\varepsilon''(\omega)\,d\omega}{\omega-\omega_1}$$

$$=\mathrm{P}\int_{0}^{\infty}\varepsilon''(\omega)\,d\omega\left[\frac{1}{\omega-\omega_1}+\frac{1}{\omega+\omega_1}\right]$$

where in the first integral ω has been replaced by $-\omega$ and the integration limits have been interchanged. Finally we obtain

$$\varepsilon'(\omega_1)-1=\frac{2}{\pi}\,\mathrm{P}\int_{0}^{\infty}\frac{\varepsilon''(\omega)\,\omega\,d\omega}{\omega-\omega_1} \tag{36.29}$$

This integral is taken only over positive frequency values. In particular, for the static dielectric constant we have

$$\varepsilon'(0) - 1 = \frac{2}{\pi} \int_0^\infty \frac{\varepsilon''(\omega)}{\omega}\, d\omega \tag{36.30}$$

Absorption usually takes place in a restricted frequency band. But as was shown, the imaginary part of $\varepsilon(\omega)$, that is, $\varepsilon''(\omega)$, is responsible for absorption. Therefore, the integral (36.29) is actually taken only over the absorption band and yields the real part of the dielectric constant for all frequencies ω_1 in the form

$$\varepsilon(\omega_1) - 1 = \frac{\text{constant}}{\overline{\omega^2} - \omega_1^2} \tag{36.31}$$

where $\overline{\omega^2}$ is the mean square of the frequency for the high-absorption band, and ω_1 lies outside the high-absorption band.

If $\varepsilon'(\omega)$ and $\varepsilon''(\omega)$ are obtained by measurement, its correctness is verified by checking the integral relationship (36.29).

The significance of dispersion relations goes far beyond the electrodynamics of continuous media. In the physics of elementary particles there occur similar relationships between the amplitudes of elastic and inelastic scattering, which like the Kramers-Kronig relations express the causality principle. The validity of the causality principle for distances of 10^{-14} cm and less has been repeatedly challenged, and experimental verification of dispersion relations here is of great fundamental interest.

EXERCISES

1. Express the total number of electrons in a unit volume in terms of the integral of the imaginary part of the dielectric constant.

Solution. Assuming that ω_1 is very large, from (36.23) we get $\varepsilon(\omega_1) = 1 - 4\pi Ne^2/(m\omega_1^2)$. Further, since the frequency interval in which the absorption is high belongs wholly to the domain in which $\omega \ll \omega_1$,

$$\varepsilon'(\omega_1) - 1 = -\frac{4\pi Ne^2}{m\omega_1^2} = -\frac{2}{\pi\omega_1^2} \int_0^\infty \varepsilon''(\omega)\, \omega\, d\omega$$

whence

$$N = \frac{m}{2\pi^2 e^2} \int_0^\infty \varepsilon''(\omega)\, \omega\, d\omega$$

2. Deduce the dispersion relations for metals.

Solution. The difference between metals and dielectric substances is that in the former $\varepsilon''(0)$ becomes infinite, like $4\pi\sigma/\omega$ as $\omega \to 0$. Therefore for metals the function

$$\varepsilon - \frac{4\pi i\sigma}{\omega}$$

possesses the same properties as ε for dielectric materials. The principal value of $\int_{-\infty}^{\infty} d\omega/\omega$ is zero, as was shown before in general form. The integral of the term involving conductivity along the small semicircle yields $-4\pi^2\sigma/\omega_1$. This quantity should be added to the right-hand side of (36.27).

37

THEORY OF DISPERSION

Classical Theory of Dispersion. The first explanation of dispersion was offered on the basis of the classical electron theory. It is based on the model concept of quasi-elastically bound electrons possessing a certain natural oscillation frequency ω_0. The equation of forced oscillations of such electrons in an electric field $\mathbf{E} = \mathbf{E}_0 e^{-i\omega t}$ has the form

$$m(\ddot{\mathbf{r}} + \omega_0^2\mathbf{r}) = e\mathbf{E}_0 e^{-i\omega t} \tag{37.1}$$

(see the supplementary problem to [Sec. 7] at the end of Volume 1). At $\omega \neq \omega_0$ the solution involving the frequency of the external field can be written as

$$\mathbf{r} = \frac{e}{m(\omega_0^2 - \omega^2)}\mathbf{E} \tag{37.2}$$

The expression for the dielectric constant is obtained from this in the same way as from Eq. (36.23) for the case of high frequencies:

$$\varepsilon(\omega) = 1 + \frac{4\pi Ne^2}{m(\omega_0^2 - \omega^2)} \tag{37.3}$$

In the more general case, when a medium contains electrons with different natural oscillation frequencies, the dispersion formula for N_i electrons per unit volume has the form

$$\varepsilon(\omega) = 1 + \sum_i \frac{4\pi N_i e^2}{m(\omega_{0i}^2 - \omega^2)} \tag{37.4}$$

At high frequencies, when

$$\omega \gg \omega_{0i}$$

for all i's, Eq. (37.4) transforms into the asymptotic form (36.24). The analytical form of the dependence (37.3) coincides with that yielded by Eq. (36.31) if ω is substituted for ω_1 and ω_0^2 for $\overline{\omega^2}$. The latter relationship indicates that electromagnetic oscillations in a medium are subjected to great attenuation when their frequency is close to the natural oscillation frequency of the electrons in the medium.

Owing to its correspondence to Eq. (36.21), actually derived from the causality postulate only, at frequencies lying outside the strong absorption band the classical dispersion formula beyond doubt has the correct analytical form. Nor has it lost its meaning at present.

When the quantum theory of atomic structure was being elaborated, it became apparent that electrons in a medium are not in any way bound quasi-elastically. Thus, in the old quantum theory of Bohr the electrons were supposed to travel around the nuclei along certain stable orbits. But if this were really so the dispersion formula (37.4) would involve the frequencies of the electrons' rotation about the nuclei. Actually it involves not these frequencies but the characteristic frequencies of quantum transitions, which are, according to Bohr's condition, determined by the energy differences of the electrons on the orbits: $\omega_{12} = (E_1 - E_2)/h$.

It is impossible to obtain this from the classical picture of motion, even taking Bohr's postulates into account. That is why H.A. Kramers and W. Heisenberg suggested another, more abstract, derivation of the dispersion formula so that it would involve precisely the transition frequencies. But it is apparent from the derivation of Eq. (37.3) that there should, nevertheless, exist certain quantities that vary with time according to a harmonic law of the type (37.1). Heisenberg analyzed what actually varies harmonically and obtained the equations of motion for the matrix elements of [27.13], thereby arriving at the matrix representation of quantum mechanical equations independently from E. Schrödinger. The equivalence of both concepts was demonstrated somewhat later.

The Wave Equation for an Atom in a Given Radiation Field. We shall now derive the quantum dispersion formula, which, as we have said, is in a form similar to the classical formula. The calculation scheme is also in many ways similar to the classical: first the mean dipole moment induced by the field is calculated; then the polarization is determined from the dipole moment; finally, knowing the polarization, $\varepsilon(\omega)$ is found. To determine the dipole moment we must first, in accordance with the general rules of quantum mechanics, determine the wave function of an atom excited by the field of a light wave, and then perform quantum mechanical averaging.

It is convenient here to introduce the external field's time dependence in real form so as not to substitute a non-Hermitian complex operator into the Schrödinger equation. We put

$$\mathbf{E} = \mathbf{E}_0 \cos \omega t \tag{37.5}$$

We shall assume the radiation wavelength to be great in comparison with the dimensions of the atom. For visible light, for example, it is 10^4 times greater than the atomic radius. Therefore the field $\mathbf{E}_0$ is virtually uniform and in the same phase over the whole atom. It is apparent from this that the correction to the Hamiltonian, or the perturbation energy, by [33.53] has the form

$$\hat{\mathscr{H}}^{(1)} = -(\hat{\mathbf{d}} \cdot \hat{\mathbf{E}}) \tag{37.6}$$

where we have used $\hat{\mathscr{H}}^{(1)}$ instead of $\hat{V}$, and $\hat{\mathbf{d}}$ is the operator of the dipole moment of the atom. Denoting the Hamiltonian of the unperturbed system as $\hat{\mathscr{H}}^{(0)}$, we find that the Schrödinger equation takes the form

$$-\frac{h}{i} \frac{\partial \psi}{\partial t} = \hat{\mathscr{H}}^{(0)} \psi + \hat{\mathscr{H}}^{(1)} \psi \tag{37.7}$$

Separating the wave function into the unperturbed part $\psi^{(0)}$ and the perturbation function $\psi^{(1)}$, which is linear with respect to $\mathbf{E}$ and therefore small, we obtain an equation used before (see [32.29]):

$$-\frac{h}{i} \frac{\partial \psi^{(1)}}{\partial t} - \hat{\mathscr{H}}^{(0)} \psi^{(1)} = \hat{\mathscr{H}}^{(1)} \psi^{(0)} \tag{37.8}$$

The perturbation function $\psi^{(1)}$ is represented in the form of a series expansion over the unperturbed wave functions with time dependent coefficients:

$$\psi^{(1)} = \sum_n c_n (t)\, \psi_n^{(0)} \tag{37.9}$$

For the expansion coefficients we obtain a set of ordinary differential equations [32.32] in which the amplitude c_0 of the unperturbed state $\psi_0^{(0)}$ in the right-hand side has been replaced by unity:

$$-\frac{h}{i} \frac{dc_n}{dt} = \int (\psi_n^0)^* \hat{\mathscr{H}}^{(1)} \psi_0^{(0)}\, dV \tag{37.10}$$

where the time dependence of the right-hand side is known:

$$\psi_0^{(0)} \propto e^{-iE_0 t/h}, \qquad (\psi_n^{(0)})^* \propto e^{iE_n t/h}, \qquad \hat{\mathscr{H}}^{(1)} \propto \cos \omega t ,$$

Using the usual notation for matrix elements [27.3],

$$\mathbf{d}_{n0} \equiv \int (\psi_n^{(0)})^* \hat{\mathbf{d}} \psi_0^{(0)}\, dV \equiv \mathbf{d}'_{n0} e^{i(E_n - E_0)t/h}$$
$$\equiv \mathbf{d}'_{n0} e^{i\omega_{n0} t} \tag{37.11}$$

we write Eq. (37.10) as follows:

$$-\frac{h}{i}\frac{dc_n}{dt}=-\frac{1}{2}\left[e^{i(\omega_{n0}+\omega)t}+e^{i(\omega_{n0}-\omega)t}\right](\mathbf{E}_0\cdot\mathbf{d}'_{n0}) \qquad (37.12)$$

In order to integrate this equation we must impose a certain initial condition upon c_0. It is natural to suppose that the external field acts for a sufficiently long time so that none of the transition processes associated with the switching on of the field affect the state of the atom by the given time. We can assume, for example, that the external field depends upon time according to the law

$$\mathbf{E}=\mathbf{E}_0 e^{\alpha t}\cos\omega t, \qquad \text{for} \quad t<0, \quad \alpha>0, \quad \alpha\ll\omega$$

$$\mathbf{E}=\mathbf{E}_0\cos\omega t, \qquad \text{for} \quad t\geqslant 0 \qquad (37.13)$$

The amplitude here gradually increases to its constant value $\mathbf{E}_0$. This field-variation law must be substituted into (37.12), integration must be performed from $-\infty$ to any $t \gtrless 0$, and α must be tended to zero. After this we obtain a single dependence of c_n upon t for any instant:

$$c_n=\frac{1}{2h}\left(\frac{e^{i(\omega_{n0}+\omega)t}}{\omega_{n0}+\omega}+\frac{e^{i(\omega_{n0}-\omega)t}}{\omega_{n0}-\omega}\right)(\mathbf{E}_0\cdot\mathbf{d}'_{n0}) \qquad (37.14)$$

Induced Dipole Moment. The mean value of the dipole moment is calculated according to the general formula [25.19] for quantum mechanical averages:

$$\langle\mathbf{d}\rangle=\int\psi^*\hat{\mathbf{d}}\psi\,dV=\int(\psi^{(0)*}+\psi^{(1)*})\,\hat{\mathbf{d}}\,(\psi^{(0)}+\psi^{(1)})\,dV \quad (37.15)$$

The term quadratic with respect to $\psi^{(1)}$ should, of course, be dropped, since the calculations are carried up to quantities proportional to the first power of $\mathbf{E}_0$. The term $\int\psi^{(0)*}\hat{\mathbf{d}}\psi^{(0)}\,dV$ bears no relation to the question of the polarization caused by the external field. Besides, if the function $\psi^{(0)}$ is odd or even with respect to inversion, then this natural dipole moment is zero, since the whole integrand is odd. Consequently, the mean dipole moment responsible for the dispersion is

$$\langle\mathbf{d}\rangle=\int(\psi^{(0)*}\hat{\mathbf{d}}\psi^{(1)}+\psi^{(1)*}\hat{\mathbf{d}}\psi^{(0)})\,dV \qquad (37.16)$$

We substitute the expansion (37.9) and integrate the series term by term to get

$$\langle\mathbf{d}\rangle=\sum_n\left(c_n\int\psi_0^{(0)*}\hat{\mathbf{d}}\psi_n^{(0)}\,dV+c_n^*\int\psi_n^{(0)*}\hat{\mathbf{d}}\psi_0^{(0)}\,dV\right) \qquad (37.17)$$

The integrals involved here are again matrix elements of the dipole moment. Substituting their expressions from (37.11), we write the

mean dipole moment as

$$\langle \mathbf{d} \rangle = \sum_n (c_n e^{-i\omega_{no}t}\, \mathbf{d}'_{0n} + c_n^* e^{i\omega_{no}t} \mathbf{d}'_{n0}) \tag{37.18}$$

With the help of Eq. (37.14) for c_n, we obtain the final expression

$$\langle \mathbf{d} \rangle = \frac{1}{2h} \sum_n \left[\left(\frac{e^{i\omega t}}{\omega_{n0}+\omega} + \frac{e^{-i\omega t}}{\omega_{n0}-\omega} \right) \mathbf{d}_{0n} (\mathbf{E}_0 \cdot \mathbf{d}_{n0}) + \left(\frac{e^{i\omega t}}{\omega_{n0}-\omega} + \frac{e^{-i\omega t}}{\omega_{n0}+\omega} \right) \mathbf{d}_{n0} (\mathbf{E}_0 \cdot \mathbf{d}_{0n}) \right] \tag{37.19}$$

Here we wrote $\mathbf{d}_{0n}$ instead of $\mathbf{d}'_{0n}$ because the time factors $e^{\pm i\omega_{no}t}$ cancel out.

In order to calculate the polarization of an atom by a light wave it is sufficient to know the projection of the dipole moment on the direction of the field. If, for example, the electric field of an incidene wave is directed along the x axis, then (37.19) involves only the components of the transition moments along the x axis, that is, the matrix element of x:

$$\langle d \rangle = \frac{e^2}{2h} \sum_n \left[\left(\frac{e^{i\omega t}}{\omega_{n0}+\omega} + \frac{e^{-i\omega t}}{\omega_{n0}-\omega} \right) E_{x0} | x_{n0} |^2 + \left(\frac{e^{i\omega t}}{\omega_{n0}-\omega} + \frac{e^{i\omega t}}{\omega_{n0}+\omega} \right) E_{x0} | x_{n0} |^2 \right] \tag{37.20}$$

In substituting $| x_{n0} |^2$ for $x_{n0}x_{0n}$ use was made of the Hermitian nature of the matrix elements ($x_{0n} = x^*_{n0}$). A simple algebraic transformation and introduction of the electric field $\mathbf{E}$ instead of its amplitude $\mathbf{E}_0$ yields

$$\langle \mathbf{d} \rangle = \sum_n \frac{2\omega_{n0} e^2 | x_{n0} |^2}{h (\omega_{n0}^2 - \omega^2)} \mathbf{E} \tag{37.21}$$

The Dispersion Formula. Substituion of the mean dipole moment into the expression for electric polarization yields the dielectric constant:

$$\varepsilon(\omega) = 1 + \frac{4\pi N e^2}{h} \sum_n \frac{2\omega_{n0} | x_{n0} |^2}{\omega_{n0}^2 - \omega^2} \tag{37.22}$$

The dependence of ε upon frequency is exactly the same as in the classical dispersion theory (37.4), but the meaning of the quantities characterizing the medium is, of course, entirely different. The frequencies ω_{n0} replace the natural oscillation frequencies of the electrons, and this fact, as was pointed out, formed the basis of the theory of Kramers and Heisenberg.

Let us now compare the numerators of the expressions (37.4) and (37.22) and show that they can be ascribed similar meaning.

First of all it is obvious that

$$N=\sum_i N_i \tag{37.23}$$

where N is the total number of electrons. Denoting the relative fraction of each oscillation as

$$f_i=\frac{N_i}{N} \tag{37.24}$$

we have

$$\sum_i f_i=1 \tag{37.25}$$

Comparing the classical and quantum dispersion formulas, we see that they become identical if we put

$$f_n\equiv\frac{2m\omega_{n0}}{h}\,|x_{n0}|^2 \tag{37.26}$$

and prove that this definition of f_n satisfies condition (37.25). In other words, we have to develop the equation

$$\sum_n \frac{2m\omega_{n0}}{h}\,|x_{n0}|^2=1 \tag{37.27}$$

We shall proceed from the commutation relation between the momentum and position operators of the electron [24.17]:

$$\hat{p}_x\hat{x}-\hat{x}\hat{p}_x=\frac{h}{i} \tag{37.28}$$

which we rewrite in matrix representation [26.14]:

$$\sum_n (p_{0n}x_{n0}-x_{0n}p_{n0})=\frac{h}{i} \tag{37.29}$$

From [27.13], the position and momentum matrix elements are connected by relationships of the type [27.25]:

$$p_{0n}=im\omega_{0n}x_{0n},\qquad p_{n0}=im\omega_{n0}x_{n0},\qquad \omega_{0n}=-\omega_{n0} \tag{37.30}$$

With this in mind, we find that

$$\sum_n (im\omega_{0n}\,|x_{0n}|^2-im\omega_{n0}\,(x_{0n}|^2)=\frac{h}{i} \tag{37.31}$$

Taking into account that $\omega_{0n}=-\omega_{n0}$, we arrive at Eq. (37.27).

It was probably the need for this equation that led Heisenberg to the idea of noncommuting coordinate and momentum matrices.

Note that the quantities f_n (called *oscillator strengths*) are proportional to the same matrix elements that are involved in the emission and absorption probabilities of the corresponding quanta

[Sec. 36]. That is why the dispersion properties of a substance are associated with the intensities of the spectral lines it emits.

Frequency Close to Resonance. The dispersion formula (37.22) becomes meaningless if the radiation frequency ω is close to one of the transition frequencies ω_{n0}, since at $\omega = \omega_{n0}$ the corresponding denominator becomes zero. This case requires special consideration.

If the frequency of a wave impinging on an atom is close to the transition frequency, "pumping" of the nth level occurs, that is, the number of atoms in that state increases. But an excited atom is capable of emitting quanta, therefore the "pumping" cannot go on indefinitely.

To take this into account we must rewrite the equations for the amplitudes of the excited states, c_n. In Eq. (37.12) it is sufficient to retain only the second term on the right. To it we must add another term describing the change in the probability amplitude c_n due to emission of quanta.

The excited state has a finite lifetime and, therefore, a line width other than zero. Accordingly, the frequency of a quantum should not be assumed in advance to be exactly equal to the frequency of the incident light. Let us call the frequency of the emitted quantum ω_k, where k stands for all the parameters determining the state of the quantum, that is, the wave vector and polarization.

We denote the matrix element corresponding to an act of emission of a quantum by the symbol $\mathscr{H}_{nk}$. Its time dependence is determined by an exponential factor, namely

$$\mathscr{H}_{nk} = e^{i(\omega_{n0}-\omega_k)t}\,\mathscr{H}'_{nk} \tag{37.32}$$

Hence the change in the amplitude c_n is given by the equation

$$-\frac{h}{i}\frac{dc_n}{dt} = -\frac{1}{2}(\mathbf{E}\cdot\mathbf{d}'_{n0})\,e^{i(\omega_{n0}-\omega)t}\sum_k e^{i(\omega_{n0}-\omega_k)t}\,\mathscr{H}'_{nk}c_k \tag{37.33}$$

The first term here corresponds to the absorption of incident radiation by the ground state in the transition to the nth state, and the second term corresponds to the emission of a quantum of frequency ω_k. Summation is carried out over all such quanta. The amplitude of the state in which a quantum of frequency ω_k is present is denoted by c_k, which, in turn, varies due to the reabsorption of these quanta. Therefore

$$-\frac{h}{i}\frac{dc_k}{dt} = \mathscr{H}'_{nk}e^{-i(\omega_{n0}-\omega)t}\,c_n \tag{37.34}$$

Here summation is not carried out, because in such reabsorption the system simply reverts to the state n. Equations (37.33) and

(37.34) are written according to the general scheme established in [Sec. 32]. The initial conditions, as usual, have the form $c_n(0) = 0$, $c_k(0) = 0$.

Let us introduce instead of c_k a new unknown function c'_k in the following way:

$$c_k = c'_k e^{i(\omega_k - \omega_{n0})t} \tag{37.35}$$

Function c'_k satisfies the equation

$$\frac{dc'_k}{dt} + i(\omega_k - \omega_{n0})\, c'_k = -\frac{i}{h}\mathscr{H}'_{kn} c_n \tag{37.36}$$

which does not involve time explicity.

We shall first consider it as a linear nonhomogeneous differential equation whose right-hand side is known. The solution satisfying the initial condition has the form

$$c'_k = -\frac{i}{h}|\mathscr{H}'_{kn} \int_0^t \exp\,[i(\omega_k - \omega_{n0})(t' - t)]\, c_n(t')\, dt' \tag{37.37}$$

This integral is conveniently transformed by parts using the initial condition for c_n. After carrying out the transformation we obtain

$$c'_k = \frac{\mathscr{H}'_{kn}}{h} \int_0^t \frac{\exp\,[i(\omega_k - \omega_{n0})(t' - t)] - 1}{\omega_k - \omega_{n0}} \frac{dc_n}{dt'}\, dt' \tag{37.38}$$

Now we substitute c'_k into Eq. (37.33) and interchange the summation over k and the integration with respect to t':

$$-\frac{h}{i}\frac{dc_n}{dt} = -\frac{1}{2}(\mathbf{E}_0 \cdot \mathbf{d}'_{n0})\, e^{i(\omega_{n0} - \omega)t}$$
$$+ \frac{1}{h}\int_0^t dt' \frac{dc_n}{dt'} \sum_k |\mathscr{H}'_{kn}|^2 \frac{\exp\,[i(\omega_k - \omega_{n0})(t' - t)] - 1}{\omega_k - \omega_{n0}} \tag{37.39}$$

The sum over k actually reduces to an integral. Denoting the number of quantum states per unit energy interval as $\rho(E_k)\, dE_k = h\rho(E_k)\, d\omega_k$, the change from summation to integration can be symbolically written as follows:

$$\sum_k \to h \int \rho(E_k)\, d\omega_k$$

Thus, Eq. (37.39) involves the integral

$$\int \rho(E_k)\, |\mathscr{H}'_{nk}|^2 \frac{\exp\,[i(\omega_k - \omega_{n0})(t' - t)] - 1}{\omega_k - \omega_{n0}}\, d\omega_k \tag{37.40}$$

If the difference $t - t'$ is great, the integral reduces to

$$-\pi i \rho(\omega_{n0})\, |\mathscr{H}'_{nk}|^2 \tag{37.41}$$

(see supplementary problem (c) to [Sec. 32] at the end of Volume 1). It can be seen from this that it is not time dependent. The resonance frequency $\omega_k = \omega_{n0}$ is substituted here into both $\rho(\omega_k)$ and the matrix element $\mathscr{H}'_{nk}$. This result is derived in the same way as the very similar formula [32.42] in the theory of quantum transitions. The substantiation for this is that the maximum of the integrand is the more acute the greater the difference $t - t'$. We thus obtain

$$\frac{dc_n}{dt} = -\frac{i}{2h}(\mathbf{E}_0 \cdot \mathbf{d}'_{n0})\, e^{i(\omega_{n0}-\omega)t} - \frac{\pi}{h}\rho(\omega_{n0})\,|\mathscr{H}'_{nk}|^2 c_n \quad (37.42)$$

The factor of c_n is, according to [Sec. 36], one-half the probability Γ of spontaneous transition from the nth state with the emission of a quantum:

$$\frac{\Gamma}{2} \equiv \frac{\pi}{h}|\mathscr{H}'_{nk}|^2 \rho(\omega_{n0}) \quad (37.43)$$

Anomalous Dispersion. We must now determine the probability amplitude c_n, and from it the induced dipole moment, for the case of resonance. The equation for c_n has, from (37.42) and (37.43), the form

$$\frac{dc_n}{dt} + \frac{\Gamma}{2}c_n = -\frac{i}{2h}(\mathbf{E}_0 \cdot \mathbf{d}'_{n0})\, e^{i(\omega_{n0}-\omega)t} \quad (37.44)$$

whence

$$c_n = -\frac{i}{2h}(\mathbf{E}_0 \cdot \mathbf{d}'_{n0})\, e^{-\Gamma t/2} \int_0^t \exp\left[\frac{\Gamma t'}{2} + i(\omega_{n0} - \omega)t'\right] dt'$$

The term due to the lower integration limit involves the exponentially decreasing factor $e^{-\Gamma t/2}$. Therefore, in integrating it is sufficient to substitute the upper limit. For the required amplitude c_n we obtain

$$c_n = \frac{1}{2h}(\mathbf{E}_0 \cdot \mathbf{d}'_{n0}) \frac{\exp[i(\omega_{n0} - \omega)t]}{\omega_{n0} - \omega - i\Gamma/2} \quad (37.45)$$

To it corresponds an induced dipole moment which, according to (37.20), has the form

$$\langle d \rangle = \frac{e^2}{h} \frac{|x_{n0}|^2}{\omega_{n0} - \omega - i\Gamma/2} \quad (37.46)$$

The expression holds only in the resonance neighbourhood, that is, at $|\omega_{n0} - \omega| \approx \Gamma/2$. In particular, it does not satisfy the parity requirements (36.14) and (36.15). But far from resonance the formula (37.27) is adequate. There is no precise and general formula applicable in both regions.

The dielectric constant near resonance is a complex function of frequency:

$$\varepsilon(\omega)=1+\frac{4\pi Ne^2}{h}\frac{|x_{n0}|^2}{\omega_{n0}-\omega-i\Gamma/2} \tag{37.47}$$

From [Sec. 36] and (37.43), Γ is also expressed in terms of $|x_{n0}|^2$:

$$\Gamma=\frac{4e^2\omega_{n0}^3}{hc^3}|x_{n0}|^2 \tag{37.48}$$

Note that at $\omega=\omega_{n0}-i\Gamma/2$ (a point in the complex half-plane lying below the real axis) $\varepsilon(\omega)$ becomes infinite. But above that plane $\varepsilon(\omega)$ is everywhere finite, as it should be in accordance with the causality principle (Sec. 36).

Separating the real and imaginary parts of (37.47), we obtain

$$\varepsilon'(\omega)=1+\frac{4\pi Ne^2}{h}\frac{(\omega_{n0}-\omega)|x_{n0}|^2}{(\omega_{n0}-\omega)^2+\Gamma^2/4} \tag{37.49}$$

$$\varepsilon''(\omega)=\frac{2\pi Ne^2}{h}\frac{|x_{n0}|^2\,\Gamma}{(\omega_{n0}-\omega)^2+\Gamma^2/4} \tag{37.50}$$

Hence, from the general requirement (Sec. 36), $\varepsilon''=0$. Note that here the result of Exercise 1, Section 36, cannot be applied, since as $\omega\to\infty$ the imaginary part $\varepsilon''(\omega)$ does not fall off fast enough. This however should not be seen as a shortcoming of Eq. (37.50), which from the deduction itself holds only near resonance.

Let us study the behaviour of the real and imaginary parts of $\varepsilon(\omega)$ as the frequency increases. Far from resonance, at $\omega<\omega_{n0}$ the real part $\varepsilon'(\omega)$ increases with ω as $(\omega_{n0}-\omega)^{-1}$. Such a dependence is called *normal dispersion*. At $\omega=\omega_{n0}-\Gamma/2$ the quantity ε' passes its maximum and then decreases to its minimum at $\omega=\omega_{n0}+\Gamma/2$; this is the domain of so-called *anomalous dispersion*. After the minimum, expression (37.49) increases in relative value and tends to unity from negative values, so that here too the dispersion is normal.

In the anomalous dispersion domain the maximum of $\varepsilon'(\omega)$ is at $\omega=\omega_{n0}$. Note that absorption displaces the resonance somewhat from $\omega=\omega_{n0}$, but here we do not consider this effect.

The normal dispersion curve can be observed directly by the *method of crossed prisms*. The first prism breaks down the image of a slot into a spectrum, the second, placed vertically, dispalces the spectrum up or down, according to the value of $\varepsilon'(\omega)$ (Figure 48). It will be shown in the next section that there is a connection between the function $\varepsilon'(\omega)$ and the refractive index of a substance.

It should be noted that the quantity Γ defines the so-called *natural width of spectral lines* associated with radiation damping. In experiments the Doppler spread of lines, due to the thermal motion of atoms, is usually strongly manifest. At the middle of a line the

Doppler spread obscures the natural width, but owing to the exponential form of the velocity distribution of atoms or molecules it

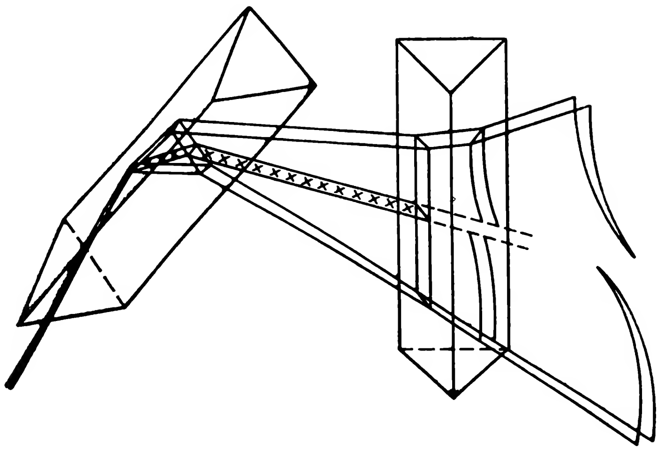


Figure 48

yields a profile of the form $\exp[-mc^2(\omega-\omega_0)^2/(2\theta\omega_0^2)]$, where m is the mass of the atom. This means that the natural spread appears at the "wings" of the line.

EXERCISES

1. Show that the static dielectric constant is always greater than unity.
Hint. From (37.22)

$$\varepsilon(0)=1+\frac{8\pi Ne^2}{h}\sum_n \frac{|x_{n0}|^2}{\omega_{n0}}$$

2. Derive the classical dispersion formula taking into account electromagnetic radiation by an electron brought into oscillation by an incident wave.

Solution. Introducing the radiation friction force [20.30] into Eq. (37.1) we write it in the form

$$m(\ddot{\mathbf{r}}+\omega_0^2\mathbf{r})=|e\mathbf{E}_0e^{-i\omega t}+\frac{2}{3}\frac{e^2}{c^3}\dddot{\mathbf{r}}$$

whence

$$\varepsilon(\omega)=1+\frac{4\pi Ne^2}{m(\omega_0^2-\omega^2)-2i\omega^3e^2/(3c^3)}$$

The imaginary term in the denominator has significant value close to $\omega = \omega_0$; therefore ω_0^3 can be substituted for ω^3 in it. But this violates the oddness condition of $\varepsilon''(\omega)$ with respect to ω (36.15).

3. Rewrite Eq. (37.47) in such a way that for a one-level system far from resonance it would transform into (37.22) and the real and imaginary parts of $\varepsilon(\omega)$ would satisfy the parity requirements (36.14) and (36.15).

Solution. Multiply the numerator and denominator (37.47) by $\omega_{n0} + \omega$. Close to resonance we can replace this quantity by $2\omega_{n0}$ in the numerator, and in the denominator replace the factor of Γ by 2ω. As a result we obtain the interpolation formula

$$\varepsilon(\omega) = 1 + \frac{8\pi N e^2}{h} \frac{\omega_{n0} |x_{n0}|^2}{\omega_{n0}^2 - \omega^2 - i\omega\Gamma}$$

38

ELECTROMAGNETIC WAVES

Plane Electromagnetic Waves. The first achievement of Maxwell's electrodynamics was the enunciation of the electromagnetic theory of light, the nature of which had till that time defied explanation. The properties of the imaginary "ether" were so strange and contradictory as to give rise to more questions in optics than it answered.

The basic equations describing the propagation of electromagnetic waves are obtained from (36.7) and (36.8). Let us write them again:

$$\text{curl}\,\mathbf{H} = -\frac{i\omega}{c}\varepsilon(\omega)\,\mathbf{E} \tag{38.1}$$

$$\text{curl}\,\mathbf{E} = \frac{i\omega}{c}\mu(\omega)\,\mathbf{H} \tag{38.2}$$

If ε and μ do not depend on position, the divergence equations are superfluous, since they follow from (38.1) and (38.2). We multiply Eq. (38.1) by $i\omega\mu(\omega)$ and substitute curl $\mathbf{E}$ for $i\omega\mu(\omega)\,\mathbf{H}$ under the curl to get

$$\text{curl}\,\text{curl}\,\mathbf{E} = -\frac{i\omega}{c}\varepsilon(\omega)\,\mathbf{E}$$

But curl curl $\mathbf{E}$ = grad div $\mathbf{E} - \nabla^2\mathbf{E} = -\nabla^2\mathbf{E}$. Hence

$$\nabla^2\mathbf{E} = -\omega^2\frac{\varepsilon\mu}{c^2}\mathbf{E} \tag{38.3}$$

We seek a solution in the form of a plane wave

$$\mathbf{E} = \mathbf{E}_0 e^{i(\mathbf{kr}-\omega t)} \tag{38.4}$$

Substitution into (38.3) yields

$$k^2 = \frac{\omega^2}{c^2} \varepsilon(\omega) \mu(\omega) \tag{38.5}$$

If ε and μ are real quantities, that is, if the frequency lies far from the anomalous dispersion region, the ratio ω/k is equal to the wave's phase velocity u [19.7]:

$$u = \frac{\omega}{k} = \frac{c}{(\varepsilon\mu)^{1/2}} \tag{38.6}$$

The group velocity of the wave is (see [19.8])

$$\mathbf{v} = \frac{\partial \omega}{\partial \mathbf{k}} \tag{38.7}$$

As was shown in [Sec. 19], this is the velocity at which a group of waves, or *wave packet*, travels. In an absorbing medium a wave packet spreads out. At high absorption this affects the dimensions of the packet itself, and Eq. (38.7) becomes meaningless.

From Eqs. (38.1) and (38.2) we obtain the relationship between the electric and magnetic fields of a plane monochromatic wave (38.4):

$$\operatorname{curl} \mathbf{H} = i\mathbf{k} \times \mathbf{H} = -\frac{i\omega\varepsilon}{c} \mathbf{E} \tag{38.8}$$

Substituting

$$\mathbf{k} = \frac{\omega}{c} (\varepsilon\mu)^{1/2} \mathbf{k}_0$$

where $\mathbf{k}_0$ is a unit vector along the direction of propagation of the wave, we obtain

$$\mathbf{k}_0 \times \mathbf{H} = -\left(\frac{\varepsilon}{\mu}\right)^{1/2} \mathbf{E} \tag{38.9}$$

or from (38.2)

$$\mathbf{k}_0 \times \mathbf{E} = \left(\frac{\mu}{\varepsilon}\right)^{1/2} \mathbf{H} \tag{38.10}$$

Thus, vectors $\mathbf{E}$ and $\mathbf{H}$ are perpendicular to $\mathbf{k}_0$ and also are mutually perpendicular, but they are not equal in absolute magnitude, as in vacuum.

The energy flux in an electromagnetic wave is determined as the time average of the Poynting vector, which according to Eq. (35.21) transposes to the form

$$\langle \mathbf{P} \rangle = \frac{c}{4\pi} \langle \mathbf{E} \times \mathbf{H} \rangle = \frac{c}{8\pi} \operatorname{Re} (\mathbf{E} \times \mathbf{H}^*)$$
$$= \frac{c}{16\pi} (\mathbf{E} \times \mathbf{H}^* + \mathbf{E}^* \times \mathbf{H})$$

In view of the perpendicularity of $\mathbf{E}$ and $\mathbf{H}$, we obtain from this

$$\langle \mathbf{P} \rangle = \frac{c}{8\pi} |\mathbf{E}| |\mathbf{H}| \mathbf{k}_0 \tag{38.11}$$

The energy density of monochromatic radiation in a transparent medium at $\varepsilon'' = 0$ is obtained by dividing the absolute value of $\langle \mathbf{P} \rangle$ by the group velocity v. This almost obvious result can also be obtained by averaging the energy with respect to time.

If absorption takes place, vector $\mathbf{k}$ is complex valued:

$$k = (\varepsilon\mu)^{1/2} \frac{\omega}{c} \equiv (n + i\varkappa) \frac{\omega}{c} \tag{38.12}$$

Here n is called the *refractive index*, and $\varkappa$ is the *absorption coefficient* of the medium at the given frequency. Note, however, that $\varkappa$ need

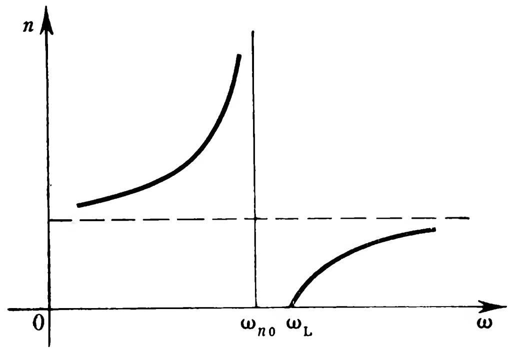


Figure 49

not be zero even at real, but negative, values of ε or μ, which they may acquire near the region of anomalous dispersion, as is apparent from Eqs. (37.3) and (37.22).

A damped wave is described by the following dependence upon position and time:

$$\mathbf{E} = \mathbf{E}_0 \exp\left(-i\omega t + in\frac{\omega}{c}x - \varkappa\frac{\omega}{c}x\right) \tag{38.13}$$

At complex-valued ε or μ vectors $\mathbf{E}$ and $\mathbf{H}$ differ in oscillation phase, as is apparent from Eqs. (38.9) and (38.10). At negative ε or μ the phase difference is equal to $|\pi/2|$.

Let $\mu = 1$ and the frequency be such that $\varepsilon(\omega) = 0$. Then the wave's electric field is longitudinal, that is, directed along $\mathbf{k}_0$. Equation (36.10) is satisfied identically. The magnetic field is in general equal to zero, because from (36.7) and (36.9) its curl and divergence are zero. Finally, Eq. (36.8) shows that curl $\mathbf{E} = 0$ and, consequently, the electric field has no transverse component.

Thus, at $\varepsilon = 0$ electric waves are longitudinal, and at $\mu = 0$ magnetic waves are longitudinal.

Figure 49 presents the dependence of the refractive index upon the frequency close to resonance at $\omega = \omega_{n0}$, disregarding damping.

At lower frequencies ($\omega < \omega_{n0}$) the relationships $\varepsilon > 1$, $n > 1$ hold. Going over to the domain $\omega > \omega_{n0}$, we find that $\varepsilon(\omega)$ is first less than zero, so that the refractive index is purely imaginary. At such frequencies waves do not propagate through a medium, damping aperiodically in space. At point $\omega = \omega_L$ the quantity $\varepsilon(\omega)$ vanishes. This point corresponds to a longitudinal electromagnetic wave. Then $\varepsilon(\omega)$ becomes greater than zero. The wave here is trans-

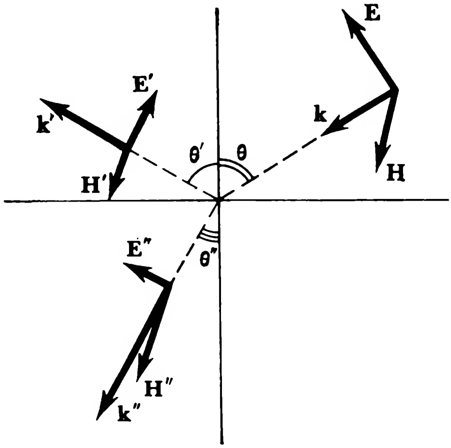


Figure 50

verse and propagates without damping, provided $\varepsilon(\omega)$ is a real quantity. At small values of Γ, that is, when the absorption band is narrow, ω_L may lie so far from ω_{n0} that the real absorption is immaterial.

Waves at the Interface of Nonabsorbing Media. We shall examine the reflection and refraction of electromagnetic waves at the interface of two nonabsorbing media. Let the incident wave be propagating in the x,z-plane and the electric field vector be in the same plane. Then the magnetic field vector is parallel to the x, y-plane separating the two media (Figure 50). Denote the refractive indices by n_1 and n_2, and the angle between the direction of the incident wave and the normal to the plane by θ (the angle of incidence).

To satisfy the boundary conditions we must assume that the wave not only penetrates the second medium but also reflects in part from the interface. This will become clear from the subsequent calculations.

Denote the angles of reflection and refraction in Figure 50 as θ' and θ''. Since the phase change along the boundary is determined by

the factors $\exp(ik_x x)$, $\exp(ik'_x x)$, and $\exp(ik''_x x)$, the x projections of the wave vectors of all three waves, k_x, k'_x, and k''_x, must be the same. For the boundary conditions to be satisfied at all values of x the equation $k_x = k'_x = k''_x$ must be satisfied. In the same way we find that the frequencies are the same. Subsequently the frequencies of all three waves are denoted simply as ω. It is apparent from Figure 50 that $k_x = k \sin\theta$, $k'_x = k' \sin\theta'$, and $k''_x = k'' \sin\theta''$. From (38.12), at $\varkappa = 0$ it follows that $k = \omega n_1/c$, $k' = \omega n_1/c$, $k'' = \omega n_2/c$. From this we obtain the equations

$$\sin\theta = \sin\theta'$$

$$n_1 \sin\theta = n_2 \sin\theta'' \tag{38.14}$$

The first is satisfied if $\theta' = \theta$, and it expresses the well-known reflection law: the angle of reflection equals the angle of incidence. In the same way we obtain the refraction law

$$\frac{\sin\theta''}{\sin\theta} = \frac{n_1}{n_2} \tag{38.15}$$

What we must do is to determine the relationship between the amplitude of the reflected and incident waves. For this we make use of the boundary conditions (28.34) and (28.37). Assuming that $\mu = 1$, as is the case for most transparent media, we considerably simplify the final formulas. It is sufficient to apply only the boundary conditions imposed on the electric field and the displacement vector; the condition for the magnetic field is satisfied automatically because of the relationships (38.9) and (38.10).

From the normal displacement components we have

$$n_1^2 (E - E') \sin\theta = n_2^2 E'' \sin\theta'' \tag{38.16}$$

which is directly apparent from Figure 50.

The condition for the tangential components of the electric field is written as follows:

$$(E + E') \cos\theta = E'' \cos\theta'' \tag{38.17}$$

Substituting the expression for n_1^2 from the refraction law (38.15) into (38.16), we obtain

$$(E - E') \sin\theta'' = E'' \sin\theta$$

We multiply this equation by $\cos\theta''$ and Eq. (38.17) by $\sin\theta$ and subtract one from the other. The amplitude of the refracted wave is then cancelled out, leaving the relation between the amplitudes of the incident and reflected waves:

$$(E - E') \sin 2\theta'' = (E + E') \sin 2\theta$$

From this, after some simple transformations we obtain

$$\frac{E'}{E}=\frac{\tan(\theta''-\theta)}{\tan(\theta''+\theta)} \tag{38.18}$$

It is apparent from the deduction that it would have been impossible to satisfy the boundary conditions (38.16) and (38.17) without first introducing the reflected wave.

Equation (38.18) was obtained by A.I. Fresnel long before Maxwell's theory from the picture of elastic transverse oscillations in an

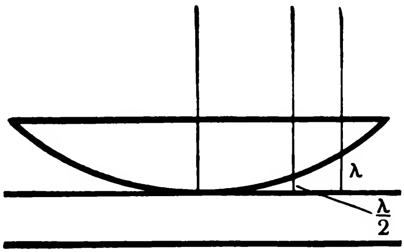


Figure 51

imaginary medium. Fresnel also obtained a similar formula for the case of a magnetic field lying in the plane of incidence. Of course, he imposed the condition on only one vector, being unaware of the other. But as pointed out, if the conditions are satisfied for one vector, they are automatically satisfied for the other.

If the direction of the incident wave approaches the normal, the electric field remains with a tangential component, written with the plus sign in Eq. (36.17) because it is in the same direction as the incident wave. But it is apparent from Fresnel's formula that the sign of the ratio E'/E depends upon what is greater, the angle of incidence or the angle of refraction. Since at small angles the sine and tangent can be replaced by the argument,

$$\lim_{\theta\to 0}\frac{E'}{E}=\lim_{\theta\to 0}\frac{\tan(\theta''-\theta)}{\tan(\theta''+\theta)}=\frac{n_1-n_2}{n_1+n_2} \tag{38.19}$$

Consequently, in the case of normal incidence on the boundary of a medium with a high refractive index the sign of the field in the reflected wave changes.

This is easily observed in the so-called Newton's rings (Figure 51). A planoconvex lens lies on a glass plate. Viewed in reflected light, a system of light and dark concentric rings is seen, according to whether the path difference of the rays reflected from the inner surface of the lens and from the plane is equal to an even or odd number

of half-waves. But the centre is dark due to a phase change of π on reflection from the external surface of the glass.

If in Eq. (38.18) $\theta + \theta'' = \pi/2$, the tangent in the denominator becomes infinite, and the amplitude of the reflected beam is zero. The corresponding angle of incidence is determined from the equation

$$\cos\theta = \cos\left(\frac{\pi}{2} - \theta''\right) = \sin\theta''$$

Substituting $\sin\theta'' = \frac{n_1}{n_2}\sin\theta$, we obtain

$$\tan\theta = \frac{n_2}{n_1} \tag{38.20}$$

(*Brewster's law*). If the magnetic field lies in the plane of incidence, E'/E does not become zero at any value of θ (Exercise 3).

Let an arbitrarily polarized wave fall on an interface. Its electric field can be resolved into two components: one whose electric field lies in the plane of incidence, and another whose electric field lies on the interface. It is only the second component that is reflected. As a consequence the reflected beam is plane-polarized. This can be detected by placing a second glass plate (not a mirror!) so that the field of the wave reflected from the first glass is now in the plane of incidence, and the angle of incidence is again equal to the value determined by Eq. (38.20). Now reflection occurs. The first and second plates are called, respectively, the polarizer and analyzer, according to the part they play in polarizing a wave and detecting that polarization.

Surface Impedance. A metal can be treated as a medium for which the imaginary part of the dielectric constant is very large. At small frequencies it is equal to $4\pi\sigma/\omega$ (see Eq. (36.24)). Using the complex dielectric constant, we can obtain equations that describe reflection from metal surfaces in the same way as Fresnel's formulas were obtained.

There exists an approximate but simpler approach to the problem, suggested by M.A. Leontovich. Instead of considering an electromagnetic wave rapidly damped in a metal we make use of the ratio between the tangential components of the field at the surface of the metal. As was shown in Section 35, a rapidly variable field penetrates a metal to a depth in inverse proportion to the square root of frequency and electrical conductivity. Consequently, the field derivatives are especially large in the direction perpendicular to the surface of the metal. But then it is apparent from Eqs. (38.1) and (38.2) that the tangential components of the field are large in comparison with the normal components. For example, if the z axis is

directed along the normal, the field has large projections

$$E_x \sim \frac{\partial H_y}{\partial z}, \qquad H_y \sim \frac{\partial E_x}{\partial z}$$

The tangential components $\mathbf{E}_t$ and $\mathbf{H}_t$ within the metal are linked by the relationship (38.9). But since they are continuous on the surface, the tangential components outside the metal satisfy the same dependence:

$$\mathbf{E}_t = \left(\frac{\mu}{\varepsilon}\right)^{1/2} \mathbf{H}_t \times \mathbf{k}_0 \tag{38.21}$$

where μ and ε refer to the region inside the metal. Thereby the reflection of an electromagnetic wave is described with the help of one complex constant

$$\zeta \equiv \left(\frac{\mu}{\varepsilon}\right)^{1/2} \tag{38.22}$$

and the boundary condition (38.21). The quantity ζ is called *the surface impedance of a metal.*

With the help of (38.21) it is easy to solve the problem on the reflection of an electromagnetic wave from a metal. Let a wave's electric field be parallel to the boundary of the metal ($\mathbf{E}_t = \mathbf{E}$). The tangential component of the incident wave's magnetic field is $|\mathbf{H}_t| = H\cos\theta$, and for the reflected wave, as is readily apparent from a construction analogous to the one in Figure 50, $|\mathbf{H}_t'| = -H'\cos\theta$. Assuming that outside the metal $\mathbf{E} = \mathbf{H}$, we obtain from (38.21)

$$E + E' = \zeta (H - H') \cos\theta = \zeta (E - E') \cos\theta \tag{38.23}$$

whence we find the ratio of amplitudes

$$\frac{E'}{E} = -\frac{1 - \zeta\cos\theta}{1 + \zeta\cos\theta} \tag{38.24a}$$

The expression for ζ involves $\sqrt{\varepsilon}$ in the denominator. But the modulus of this quantity is great (due to the high conductivity σ); hence we can write to a good approximation:

$$\frac{E'}{E} = -1 + 2\zeta\cos\theta \tag{38.24b}$$

The ratio of the field amplitudes is close to unity, which is observed in reflection from a metal.

The only exception is if the electric field lies in the plane of incidence. Then instead of (38.23) we obtain

$$\zeta (H + H') = (E - E') \cos\theta \tag{38.25}$$

If the incident wave is at a grazing angle to the surface ($\theta \approx \approx \pi/2$, $\cos\theta \ll 1$), at small values of $|\zeta|$ the amplitude of the reflected wave, E', may differ considerably from E.

Leaving aside this special case, we can assert that the electric field vector tangential to a surface, that is, the sum of the incident and reflected vectors, vanishes if the surface impedance is small enough. Thus, the greater the absorption in a medium the better the reflecting qualities of its boundary: the equation $\mathbf{E}_t = 0$ assures the absence of conductivity currents in the metal.

Cavity Resonators. In generating electromagnetic oscillations with frequencies of the order of 10^{10} Hz and higher, use is made not of circuits with lumped parameters, but of cavity resonators with walls made of polished, highly conductive metals.

Assuming for simplicity that nothing fills the cavity, in Eqs. (38.1) and (38.2) we can put $\varepsilon = 1$, $\mu = 1$. Applying the curl operator to the second, we obtain

$$\text{curl curl } \mathbf{E} = \frac{i\omega}{c} \text{curl } \mathbf{H} = \frac{\omega^2}{c^2} \mathbf{E}$$

But from (38.1) div $\mathbf{E} = 0$, so we substitute $-\nabla^2 \mathbf{E}$ for curl curl $\mathbf{E}$. Consequently, the electric (and magnetic) field satisfies the equation

$$\nabla^2 \mathbf{E} + \frac{\omega^2}{c^2} \mathbf{E} = 0 \tag{38.26}$$

Notation in terms of the Laplace operator ∇^2 is useful only when dealing with Cartesian coordinates. In most cases, however, curvilinear coordinates are used, in accordance with the shape of the resonator. Therefore the curl should be treated as the differential operation defined in [11.47].

If the surface impedance is small, the boundary condition for Eq. (38.26) is $\mathbf{E}_t = 0$, that is, the electric field should not have a tangential component. Then the normal component of the Poynting vector $\mathbf{P} = (c/4\pi)\mathbf{E}_t \times \mathbf{H}_t$ vanishes, and there are no losses in the resonator's walls.

Together with the boundary condition $\mathbf{E}_t = 0$, Eq. (38.26) represents an eigenvalue problem concerning the frequency ω similar to the eigenvalue problem in quantum mechanics [Sec. 28]. Solutions are classified according to the number of nodal surfaces on which the components of the electric vector become zero. To each solution (or *oscillation mode*, as it is called in radio engineering) there corresponds a natural frequency. If a cavity possesses a known symmetry, spherical, for example, the same frequency may correspond to several modes. In quantum mechanics this is called *degeneracy*. Examples of cavity resonators are presented in Exercises 4 and 5.

Waveguides. A *waveguide* is a long (infinite) cylindrical cavity with metallic walls. Electromagnetic waves propagate along it without scattering in space. There are two types of waveguides: electric (**E**), for which the projection of the electric field on the tube's axis is not zero, and magnetic (**H**), with a nonzero axial projection of the magnetic field.

Coaxial Waveguides. Consider a waveguide of circular cross section in which neither the electric nor the magnetic field has an axial component. We look for the field of the electromagnetic wave in the form

$$E_r = E_{0r}(r)\, e^{i(kz-\omega t)}, \qquad E_z = E_\varphi = 0 \tag{38.27a}$$

$$H_\varphi = H_{0\varphi}(r)\, e^{i(kz-\omega t)}, \qquad H_r = H_z = 0 \tag{38.27b}$$

The field equations are

$$\operatorname{div} \mathbf{E} = \frac{1}{r}\frac{\partial}{\partial r} rE_{0r} = 0 \tag{38.28}$$

$$\operatorname{curl}_\varphi \mathbf{E} = -\frac{\partial E_r}{\partial z} = -ikE_r = \frac{i\omega}{c} H_\varphi \tag{38.29}$$

$$\operatorname{curl}_r \mathbf{H} = \frac{\partial H_\varphi}{\partial z} = ikH_\varphi = -\frac{i\omega}{c} E_r \tag{38.30}$$

The other Maxwell's equations are satisfied identically.

It follows from the latter two equations that $\omega = ck$, $|\mathbf{E}| = |\mathbf{H}|$, as in vacuum. The solution of (38.28) has the form E_{0r} = constant $\times$ $\times\, r^{-1}$ (as in the case of a cylindrical capacitor). A necessary condition here is for the cross section to be a ring, not a disk, since at the centre of the disk the field would become infinite, and the problem would have no solution.

A waveguide in which $\omega = ck$ must have a doubly connected cross section. It is called *coaxial*. The wave propagating through it is called the *principal wave*.

Note also that at $\omega = ck$ there is no solution for which E_φ and H_r are other than zero. The quantity E_φ is the tangential component of the electric field on the walls of the waveguide, which must be zero. But then an equation similar to (38.30) shows that the normal component of the magnetic field, H_r, is also zero on the wall. This requirement is not satisfied by the solution of an equation similar to (38.28), having the form H_r = constant $\times\, r^{-1}$.

Waves Along Wires. Closely related to the problem of the propagation of the principal wave is that on an electromagnetic wave travelling along a long conductor. We again assume that the dependence of the field upon the coordinate along the wire is the same as in a travelling wave, that is, the field is proportional to $e^{-i(t\omega - zkz)}$

and the electric field in the transverse direction depends upon the coordinates like an electrostatic field. In the initial approximation we obtain the same result as in the problem on the principal wave. In the first approximation we must take account of the finite resistance of the wire.

Let the charge per unit length of the wire be e, and the longitudinal current be I. The charge conservation law requires that

$$\frac{\partial e}{\partial t} = -\frac{\partial I}{\partial z} \tag{38.31}$$

The potential at a given point of the wire is connected with the charge by the electrostatic relationship

$$e = C\varphi \tag{38.32}$$

where C is the capacitance of a unit length of the wire.

The potential gradient along the wire is equal to the current taken with the opposite sign and multiplied by the impedance of unit length, Z (see (35.28)). But this impedance does not involve a capacitance term, since capacitance is involved in Eq. (38.32) and does not affect the ratio between the current and the longitudinal field $-\partial\varphi/\partial z$. It is, so to say, joined in parallel, not in series.

We write therefore

$$-\frac{\partial \varphi}{\partial z} = IZ = (R - i\omega\mathscr{L})\, I \tag{38.33}$$

Differentiating both parts with respect to z and substituting $-\partial e/\partial t$ for $\partial I/\partial z$ according to (38.31), we obtain

$$\frac{\partial^2 \varphi}{\partial z^2} = \frac{1}{C}\frac{\partial^2 e}{\partial z^2} = Z\frac{\partial e}{\partial t} = (R - i\omega\mathscr{L})\frac{\partial e}{\partial t} \tag{38.34}$$

At frequencies employed to transmit signals along wires the constants of the medium are practically independent of the frequency. Therefore the factor $-i\omega$ can be replaced by the derivative $\partial/\partial t$, after which frequency is no longer involved anywhere. From this we obtain the required equation (it is called the *telegrapher's equation*):

$$\frac{1}{C}\frac{\partial^2 e}{\partial z^2} - R\frac{\partial e}{\partial t} - \mathscr{L}\frac{\partial^2 e}{\partial t^2} = 0 \tag{38.35}$$

Remember that inductance here is taken in electromagnetic units. To go over to Gaussian units it must be divided by c^2.

Neglecting resistance, we find that the signal is transmitted along wire with the speed of light. This can be shown in the following way. From the general equations $E_x = H_y$ and $E_y = H_x$ there is symmetry between the electric and magnetic field. Introducing the

potential ψ, for which

$$H_x = -\frac{\partial \psi}{\partial x}, \qquad H_y = -\frac{\partial \psi}{\partial y}$$

we see that the electric and magnetic potentials together satisfy the Cauchy-Riemann equations (Sec. 15). Hence the electromagnetic field is defined with the help of one complex potential $\varphi + i\psi$. The vector lines of one coincide with the equipotential lines of the other. We find that the factor $1/C$ in the equations of electrostatics occupies the same place as $\mathcal{L}$ in the equations of magnetostatics. As they are determined by the same complex potential, $\mathcal{L}C = 1$ (where $\mathcal{L}$ and C are expressed in Gaussian units). Therefore at $R = 0$ Eq. (38.34) turns into the wave equation

$$\frac{\partial^2 e}{\partial z^2} - \frac{1}{c^2}\frac{\partial^2 e}{\partial t^2} = 0 \tag{38.36}$$

which describes the propagation of signals with the fundamental velocity c.

EXERCISES

1. Show that if the dispersion law is expressed by Eq. (37.3), the group velocity is always less than the speed of light in vacuum.

Solution. We write the expression for the inverse of v:

$$\frac{1}{v} = \frac{\partial k}{\partial \omega} = \frac{1}{c}\left(\sqrt{\varepsilon} + \omega\frac{\partial}{\partial \omega}\sqrt{\varepsilon}\right)$$

The required inequality reduces to

$$\varepsilon + \frac{\omega}{2}\frac{\partial \varepsilon}{\partial \omega} > \sqrt{\varepsilon}$$

Since $\partial\varepsilon/\partial\omega > 0$ (normal dispersion), at $\varepsilon > 1$ the required inequality is satisfied. At $0 \leqslant \varepsilon \leqslant 1$

$$\varepsilon + \frac{\omega}{2}\frac{\partial \varepsilon}{\partial \omega} = 1 + \frac{a^2\omega_0^2}{(\omega_0^2 - \omega^2)^2}$$

where $a^2 \equiv 4\pi N e^2/m$.

This quantity is greater than $\sqrt{\varepsilon} = [1 + a^2/(\omega_0^2 - \omega^2)]^{1/2}$, which is readily established by squaring both sides of the inequality. Negative values of ε are precluded since they correspond to absorption bands.

2. A wave from a medium with a higher refraction index n_1 impinges on the interface of a medium with a smaller refraction index n_2 at an angle θ such that

$$\frac{n_1}{n_2}\sin\theta > 1$$

Investigate the wave in the second medium.

Solution. This case, as is known, is called total internal reflection. Actually, however, the wave penetrates slightly into the second medium, but it falls off exponentially, so that the energy remains in the first medium.

The normal projection of the wave vector in the second medium is

$$n_2 \frac{\omega}{c} \cos\theta'' = n_2 \frac{\omega}{c} (1-\sin^2\theta'')^{1/2}$$
$$= n_2 \frac{\omega}{c} \left(1-\frac{n_1^2}{n_2^2}\sin^2\theta\right)^{1/2}$$
$$= i\frac{\omega}{c} (n_1^2 \sin^2\theta - n_2^2)^{1/2}$$

Hence, when it enters the second medium, the wave falls off according to the law

$$\exp\left[-\frac{z\omega}{c}(n_1^2\sin^2\theta - n_2^2)^{1/2}\right]$$

The tangential component of the wave vector is equal to $(n_2\omega/c)\sin\theta$. Thus, k'' has a real component along x and an imaginary one along z. But since they are mutually perpendicular, k''^2 is real and equal to $n_2^2\omega^2/c^2$, as it should be.

3. Derive Fresnel's formula for the case of a magnetic field lying in the plane of incidence.

Answer.

$$\frac{E'}{E} = \frac{\sin(\theta''-\theta)}{\sin(\theta''+\theta)}$$

4. Determine the natural frequencies and natural oscillations of the field in a resonator with perfectly reflecting walls having the shape of a parallelepiped with sides a_1, a_2, a_3.

Solution. We choose the electric field components in the form:

$$E_x = E_{0x}e^{-i\omega t}\cos\frac{\pi n_1 x}{a_1}\sin\frac{\pi n_2 y}{a_2}\sin\frac{\pi n_3 z}{a_3}$$

$$E_y = E_{0y}e^{-i\omega t}\sin\frac{\pi n_1 x}{a_1}\cos\frac{\pi n_2 y}{a_2}\sin\frac{\pi n_3 z}{a_3}$$

$$E_z = E_{0z}e^{-i\omega t}\sin\frac{\pi n_1 x}{a_1}\sin\frac{\pi n_2 y}{a_2}\cos\frac{\pi n_3 z}{a_3}$$

where n_1, n_2, and n_3 are integers, none of which is zero. These components satisfy the boundary conditions $\mathbf{E}_t = 0$ on all the walls. From the equation $\operatorname{div}\mathbf{E} = 0$ follows the relationship

$$\frac{n_1}{a_1}E_{0x} + \frac{n_2}{a_2}E_{0y} + \frac{n_3}{a_3}E_{0z} = 0$$

between the amplitudes E_{0x}, E_{0y}, and ξ^0. The frequency equation has the form

$$\frac{\omega^2}{c^2} = \pi^2\left[\left(\frac{n_1}{a_1}\right)^2 + \left(\frac{n_2}{a_2}\right)^2 + \left(\frac{n_3}{a_3}\right)^2\right]$$

At $n_1 = n_2 = n_3 = 1$ we obtain the smallest, that is, the fundamental frequency.

5. In a spherical cavity of radius a the only nonzero component of the magnetic field is the azimuthal component, independent of the azimuth and without zeros through the polar angle ϑ at $0 < \vartheta < \pi$. Determine the oscillation frequency (the fundamental frequency for this case).

Solution. The only nonzero components are

$$\operatorname{curl}_r \mathbf{H} = \frac{1}{r \sin\vartheta} \frac{\partial}{\partial\vartheta} H_\varphi \sin\vartheta, \qquad \operatorname{curl}_\vartheta \mathbf{H} = -\frac{1}{r} \frac{\partial}{\partial r} rH_\varphi$$

The quantity curl curl $\mathbf{H}$ has a component along the azimuth φ:

$$\operatorname{curl}_\varphi \operatorname{curl} \mathbf{H} = -\left(\frac{1}{r} \frac{\partial^2}{\partial r^2} rH_\varphi + \frac{1}{r^2} \frac{\partial}{\partial\vartheta} \frac{1}{\sin\vartheta} \frac{\partial}{\partial\vartheta} H_\varphi \sin\vartheta\right)$$

$$= -\frac{\omega^2}{c^2} H_\varphi$$

The spherical function of the lowest order with no zeros in the domain $0 < < \vartheta < \pi$ is $\sin\vartheta$. This is easily verified with the help of [29.9]. Substituting $H_\varphi = H(r) \sin\vartheta$, we obtain an equation for $H(r)$:

$$\frac{1}{r} \frac{d^2}{dr^2} rH - \frac{2H}{r} = -\frac{\omega^2}{c^2} H \equiv -k^2 H$$

It is satisfied by the function

$$H = \frac{1}{r}\left(\cos kr - \frac{\sin kr}{r}\right)$$

which is easily verified by substitution; it is regular at $r = 0$.

The electric field component E_ϑ, which is proportional to $\operatorname{curl}_\vartheta \mathbf{H}$, is zero on the surface of the sphere. Therefore at $r = a$

$$\left(\frac{d}{dr} rH\right)_{r=a} = 0$$

or

$$\cot ka = \frac{1}{ka} - ka$$

Hence $ka = 2.74$, $\omega = 2.74\, c/a$.

If we take a spherical function of a higher order, Y_l^l $(l > 1)$, the order at which the function vanishes at $\vartheta = 0$ and $\vartheta = \pi$ will be greater. This yields a higher natural frequency.

6. Determine the minimum frequency of an E-wave which can propagate through a waveguide of rectangular cross section with sides a_1, a_2.

Solution. The electric field should be taken in the form

$$E_x = E_{0x} \exp(-i\omega t + ik_z z) \cos\frac{\pi x n_1}{a_1} \sin\frac{\pi y n_2}{a_2}$$

$$E_y = E_{0y} \exp(-i\omega t + ik_z z) \sin\frac{\pi x n_1}{a_1} \cos\frac{\pi y n_2}{a_2}$$

$$E_z = E_{0z} \exp(-i\omega t + ik_z z) \sin\frac{\pi x n_1}{a_1} \sin\frac{\pi y n_2}{a_2}$$

From this we determine the oscillation frequency:

$$\omega^2 = c^2 (k_z^2 + \pi^2 n_1^2 a_1^{-2} + \pi^2 n_2^2 a_2^{-2})$$

The dependence of ω on k_z leads to dispersion, that is, the dependence of the velocity of the signal, ω/k_z, upon the frequency, ω. The minimum frequency of waves carried by the waveguide is

$$\omega = \pi c (a_1^{-2} + a_2^{-2})^{1/2}$$

The existence of a minimum frequency of the carried signal is a common property of all hollow waveguides.

There is no dispersion only in a coaxial waveguide, for which $\omega = ck$. The principal wave can propagate along it.

7. Write the formulas for the refraction index and absorption coefficient if $\varepsilon = \varepsilon' + i\varepsilon''$.

Solution. From the definition $n + i\varkappa = (\varepsilon' + i\varepsilon'')^{1/2}$ we obtain

$$n^2 - \varkappa^2 = \varepsilon', \qquad n\varkappa = \varepsilon''\varkappa/2$$

It follows from this that n^2 and $-\varkappa^2$ are the roots of the quadratic equation

$$x^2 - x\varepsilon' - \varepsilon''^2/4 = 0$$

Therefore

$$n = \left[\frac{1}{2}\varepsilon' + \left(\frac{1}{4}\varepsilon'^2 + \frac{1}{4}\varepsilon''^2\right)^{1/2}\right]^{1/2}$$

$$\varkappa = \left[\left(\frac{1}{4}\varepsilon'^2 + \frac{1}{4}\varepsilon''^2\right)^{1/2} - \frac{1}{2}\varepsilon'\right]^{1/2}$$

8. Show that at $\mu = 1$ the imaginary part of the surface impedance is negative.

Solution. From the definition of the quantity ζ we obtain

$$\zeta = \left(\frac{\mu}{\varepsilon}\right)^{1/2} = \left(\frac{1}{\varepsilon' + i\varepsilon''}\right)^{1/2} = \left(\frac{\varepsilon' - i\varepsilon''}{\varepsilon'^2 + \varepsilon''^2}\right)^{1/2} \equiv \zeta' + i\zeta''$$

We choose the root with the plus sign; since $\zeta' > 0$, the energy must be absorbed in the metal; $\zeta'' < 0$ since $\varepsilon'' > 0$.

39

SOME APPLICATIONS OF THE ELECTRODYNAMICS OF RAPIDLY VARIABLE FIELDS

Magnetic Rotation of the Polarization Plane. M. Faraday was the first to surmise that there is a connection between electromagnetic and light phenomena. It was in his quest of this connection that

he made his discovery, in 1845, of the rotation of the plane of polarization of light in a substance placed in a magnetic field.

The arrangement of the experiment is as follows. A beam of plane-polarized light is passed through a hole drilled in the core of an electromagnet, which means that part of its path is parallel to the lines of the magnetic field. The polarization plane turns through an angle proportional to the intensity of the field and the length of the path through it.

The Faraday effect is explained on the basis of Larmor's theorem [17.29], according to which a system of charges placed in a constant magnetic field H comes into uniform rotation with an angular velocity.

$$\omega_L = \frac{eH}{2mc} \tag{39.1}$$

Let us now consider a plane-polarized wave travelling along the z axis parallel to the magnetic field lines. In the absence of a magnetic field a plane-polarized wave identically resolves into a sum of two circularly polarized waves [Sec. 18]:

$$\mathbf{E} = \frac{1}{2}[(\mathbf{E}_1 + i\mathbf{E}_2) + (\mathbf{E}_1 - i\mathbf{E}_2)] \exp\left[-i\omega t + in(\omega)\frac{\omega}{c}z\right] \tag{39.2}$$

Here, $|\mathbf{E}_1| = |\mathbf{E}_2|$, $\mathbf{E}_1\mathbf{E}_2 = 0$.

Each term in the right-hand side of the equation describes a circularly polarized wave; the polarization vector of such a wave rotates about the direction of propagation with an angular velocity ω.

In a magnetic field directed along the z axis Larmor's rotation of electrons in the molecules of the medium is superimposed on this angular velocity. More exactly, going over to a reference frame connected with the molecule, acting on the electrons are circularly polarized waves of frequency $\omega \pm \omega_L$, depending on whether the polarization vector of the wave is rotating in the same or opposite direction as Larmor's procession. Correspondingly, the expression (39.2) must be changed to

$$\mathbf{E} = \frac{1}{2}\left\{(\mathbf{E}_1 + i\mathbf{E}_2)\exp\left[-i\omega t + in(\omega+\omega_L)\frac{\omega}{c}z\right] + (\mathbf{E}_1 - i\mathbf{E}_2)\exp\left[-i\omega t + in(\omega-\omega_L)\frac{\omega}{c}z\right]\right\} \tag{39.3}$$

If the field is not very strong, the refraction index can be expanded into a series:

$$n(\omega \pm \omega_L) \approx n(\omega) \pm \omega_L \frac{\partial n}{\partial \omega} \tag{39.4}$$

Then the electromagnetic wave is represented as follows:

$$\mathbf{E} = \frac{1}{2} ae^{b} \left[(\mathbf{E}_1 + i\mathbf{E}_2) e^{i\omega\omega_L/c} + (\mathbf{E}_1 - i\mathbf{E}_2) e^{-i\omega\omega_L/c} \right]$$

$$= a \left[\mathbf{E}_1 \cos\left(\frac{\omega\omega_L b}{c}\right) - \mathbf{E}_2 \sin\left(\frac{\omega\omega_L b}{c}\right) \right] \tag{39.5}$$

where $a = \exp\left[-i\omega t + in(\omega)\omega z/c\right]$, and $b = z\,(\partial n/\partial\omega)$. The second factor in (39.5) describes the uniform rotation of the wave polarization vector, the angular velocity of which is

$$\omega' = \omega_L \omega \frac{\partial n}{\partial \omega} \tag{39.6}$$

The effect is the greater the stronger the dispersion.

Natural Optical Activity. The breakdown of mirror symmetry in the domain of elementary interactions was discovered not so long ago, in 1956, by Lee and Yang. A similar asymmetry in molecules (*stereoisomerism*) has been known for more than a hundred years, since Pasteur's time. Biosynthesis processes involve stereoisomeric molecules that produce new organic stereoisomers from inorganic substances (N_2, O_2, CO_2, H_2O, and others). It is extremely difficult to understand whether the predominance of right-hand over left-hand isomers in some cases, and of left- over right-hand in others, is due to some fortuitous development in the very process of initial germination of life or whether it possesses some other, deeper meaning. The latter, however, seems extremely improbable, because weak elementary interactions, which are a thousand million times weaker than electromagnetic interactions, can hardly affect chemico-biological processes which are, in the final analysis, due to forces of an electromagnetic nature. And the laws of electromagnetism, or Maxwell's equations, are quite symmetrical with respect to transitions from right- to left-hand coordinate system (inversion).

Inversion is, as is known, a discrete operation quite independent of rotations of a coordinate system [Sec. 15]. The determinant of an inversion transformation (that is, $x \to -x$, $y \to -y$, $z \to -z$) is -1, while the determinant of continuous rotation transformations is $+1$. We can therefore picture ourselves a medium quite isotropic with respect to rotations and asymmetrical with respect to inversion. An example of such a medium is an aqueous solution of sugar: its molecules are arbitrarily oriented, but they retain their stereoisomeric property in water. A solution of "right-hand" molecules differs from a solution of "left-hand" molecules: sometimes only one of these solutions may be sweet, owing to stereoisomerism of taste receptors. Crystalline sugar is, of course, not only stereoisomeric but, like any crystal, anisotropic as well.

The propagation of electromagnetic waves in an isotropic medium containing stereoisomeric molecules possesses certain features, which will be discussed here. It was pointed out in the introduction to Section 36 that the displacement vector may be connected with the electric field in a medium not only at a given point in space but within the neighbourhood of the point as well. In the first approximation this dependence should be of tensor form:

$$D_i = \varepsilon_{ik} E_k + \alpha_{ikl} \frac{\partial E_l}{\partial x_k} \tag{39.7}$$

It is assumed here that electric displacement is a linear function not only of the field but also of its first derivative with respect to the coordinate.

In an isotropic medium a symmetric tensor of rank 2 becomes a scalar ε; from dispersion theory it can be seen that the dielectric-constant tensor is symmetric not only in electrostatics but also in the general case of a high-frequency alternating field, though in Section 37 we omitted the question for the time being (see Sec. 40). As for the rank 3 tensor α_{ikl}, in an isotropic medium it can be proportional only to an invariant tensor of rank 3 that retains its form in all rotations of the coordinate system.

The only tensor of this form was developed in [Sec. 11]; it is a completely antisymmetric tensor of rank 3 ε_{ikl}, given by the equations

$$\varepsilon_{123} = \varepsilon_{312} = \varepsilon_{231} = -\varepsilon_{213} = -\varepsilon_{132} = -\varepsilon_{321} = 1$$

When any two indices coincide, the components are zero.

Thus, in an isotropic medium $\alpha_{ikl} = \alpha\varepsilon_{ikl}$, and Eq. (39.7) takes the form

$$D_i = \varepsilon E_i + \alpha \varepsilon_{ikl} \frac{\partial E_l}{\partial x_k} \tag{39.8}$$

It is easily reduced to vector form. For $i = 1$, for example, we have $(x_1 = x,\ x_2 = y,\ x_3 = z)$:

$$D_x = \varepsilon E_x + \alpha \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} \right) \tag{39.9}$$

or, in general form,

$$\mathbf{D} = \varepsilon \mathbf{E} + \alpha \operatorname{curl} \mathbf{E} \tag{39.10}$$

In an inversion of the coordinate system vectors **D** and **E** reverse their directions; curl **E** does not change its sign since curl $\mathbf{E} = \nabla \times \mathbf{E}$, and the components of ∇, naturally, also change their signs.

Hence, Eq. (39.10) cannot be realized in a medium that is symmetric with respect to inversion of the coordinate system but it

can hold in an asymmetrical medium of the type of a sugar solution. In such a medium Maxwell's equations (38.1) and (38.2) at $\mu = 1$ take the form

$$\operatorname{curl} \mathbf{H} = -i\varepsilon \frac{\omega}{c} \mathbf{E} - \alpha \frac{i\omega}{c} \operatorname{curl} \mathbf{E} \tag{39.11}$$

$$\operatorname{curl} \mathbf{E} = \frac{i\omega}{c} \mathbf{H} \tag{39.12}$$

We seek a solution in the form of a travelling plane wave, that is, proportional to $e^{i\mathbf{kr}}$. Multiplying (39.11) by $i\omega/c$ and substituting curl $\mathbf{E}$ for $i\omega\mathbf{H}/c$, we obtain

$$\operatorname{curl}\operatorname{curl} \mathbf{E} = \varepsilon \frac{\omega^2}{i^2} \mathbf{E} + i\alpha \frac{\omega^2}{c^2} \mathbf{k} \times \mathbf{E}$$

Since it follows from (39.11) that div $\mathbf{E} = 0$, the curl curl operator can be replaced by ∇^2, and the equation for $\mathbf{E}$ reduced to

$$\left(k^2 - \varepsilon \frac{\omega^2}{c^2}\right) \mathbf{E} = i\alpha \frac{\omega^2}{c^2} \mathbf{k} \times \mathbf{E} \tag{39.13}$$

Assuming the wave to be propagating along the z axis ($k_z = k$, $k_x = k_y = 0$), we represent Eq. (39.13) in terms of the components of $\mathbf{E}$:

$$\left(k^2 - \varepsilon \frac{\omega^2}{c^2}\right) E_x - \frac{i\omega^2}{c^2} \alpha k E_y = 0$$

$$i \frac{\omega^2}{c^2} \alpha k E_x + \left(k^2 - \varepsilon \frac{\omega^2}{c^2}\right) E_y = 0 \tag{39.14}$$

For this equation to have a solution its determinant must be equal to zero:

$$\begin{vmatrix} k^2 - \varepsilon \frac{\omega^2}{c^2} & -i \frac{\omega^2}{c^2} \alpha k \\ i \frac{\omega^2}{c^2} \alpha k & k^2 - \varepsilon \frac{\omega^2}{c^2} \end{vmatrix} = 0$$

or

$$\left(k^2 - \varepsilon \frac{\omega^2}{c^2}\right)^2 - \alpha^2 k^2 \frac{\omega^4}{c^4} = 0 \tag{39.15}$$

Since the effect being considered is due to molecular asymmetry, the maximum possible value of α is not greater than the dimension of a molecule. Then the ratio of the two terms in the right-hand side of (39.10) is equal, as to order of magnitude, to the ratio of the molecular dimensions to the length of a light wave, that is, $\approx 10^{-4}$. But actually, for sugar α is considerably smaller.[6] Proceeding from

[6] The model theory shows that α depends on the interaction of asymmetrically located groups in the molecule. This reduces α by two or three orders.

this evaluation, and taking into account that in the zero approximation $k = \varepsilon^{1/2}\omega/c$, the product αk can be replaced by $\alpha\varepsilon^{1/2}\omega/c$. In the next approximation

$$k_{\pm} = \frac{\omega}{c}\left(\varepsilon \pm \varepsilon^{1/2}\alpha\frac{\omega}{c}\right)^{1/2} = \frac{\omega}{c}\varepsilon^{1/2}\left(1 \pm \frac{\alpha\omega}{2c\varepsilon^{1/2}}\right)$$
$$= \frac{\omega}{c}\varepsilon^{1/2} \pm \frac{\omega^2}{2c^2}\alpha \qquad (39.16)$$

Substituting this into expression (39.14), we obtain (in the same approximation)

$$E_x = \mp iE_y \qquad (39.17)$$

If a plane-polarized electromagnetic wave enters a medium for which $\alpha \neq 0$ (an *optically active medium*), an effect occurs closely resembling the magnetic rotation of the polarization plane. Representing a plane wave as the sum of two circularly polarized waves, we observe that in accordance with the two signs in Eq. (39.17) they have different values of k; namely, k_+ and k_-. Along a path z the resultant polarization vector turns through an angle

$$z(k_+ - k_-) = z\alpha\frac{\omega^2}{c^2} \qquad (39.18)$$

If α is not zero due to asymmetrical molecules in the solution, the concentration of the solution can be determined by measuring the rotation angle of the polarization plane. In the case of a sugar solution this would be difficult to determine by evaporation.

Cerenkov Radiation. In 1936, P. A. Cerenkov observed the passage of fast electrons through transparent media. Quite unexpectedly he observed a weak glow.

Much earlier, even before the enunciation of the theory of relativity, A. Sommerfeld theoretically examined the problem of an electron travelling faster than light and showed that electromagnetic radiation should occur similar to the sound effects in a gas through which a body is moving at supersonic velocity. Formally, both effects are of the same origin.

Interest in Sommerfeld's work naturally diminished when it was found that in vacuum nothing can travel faster than light. But I.E. Tamm and I.M. Frank noted that in a transparent medium relativistic particles can travel faster than light, the speed of light being c/n at $n > 1$. They thus explained the Cerenkov radiation as emission of light by an electron moving in a medium at a speed exceeding the phase velocity of electromagnetic waves. Radiation occurs only at a frequency for which $v > c/n(\omega)$, where v is the speed of the electron.

It should be noted that energy losses by the electrons need not be taken into account explicitly. The "faster-than-light" motion itself is sufficient for the formation of electromagnetic waves in the medium for purely kinematic reasons. Although in actual fact the electron is the energy source, the radiation intensity is determined by its velocity, and not its acceleration, as in vacuum.

The Field of an Electron Moving in a Medium. Let us determine the radiation field produced by an electron. Knowing the field, we can easily determine the retarding force acting on the electron. Obviously, the value of this force is numerically equal to the value of the energy dissipated in radiation per unit path of the electron.

To determine the field it is convenient to represent it in the form of a Fourier integral [Sec. 19]. For this we must first express the charge density and current density of the electron in that form.

Since an electron is in effect a point, its density is equal to the δ-function of the difference $\mathbf{r} - \mathbf{r}_0$ (where $\mathbf{r}_0$ is the radius vector of the point at which the electron is located at the given instant). Since it is in uniform motion, $\mathbf{r}_0 = \mathbf{v}t$. Hence the charge density ρ and current density $\mathbf{j}$ are

$$\rho = e\delta\,(\mathbf{r} - \mathbf{v}t), \qquad \mathbf{j} = e\mathbf{v}\delta\,(\mathbf{r} - \mathbf{v}t) \tag{39.19}$$

(we assume that $v_x = v$ and $v_y = v_z = 0$).

Expansion of the δ-function into a Fourier integral can be carried out according to the formulas [Sec. 26]:

$$\rho = e\delta\,(\mathbf{r} - \mathbf{v}t) = \frac{e}{(2\pi)^3}\int dk_x\, dk_y\, dk_z\, e^{i\mathbf{k}(\mathbf{r}-\mathbf{v}t)} \tag{39.20}$$

$$\mathbf{j} = \rho\mathbf{v} = \frac{e\mathbf{v}}{(2\pi)^3}\int dk_x\, dk_y\, dk_z\, e^{i\mathbf{k}(\mathbf{r}-\mathbf{v}t)} \tag{39.21}$$

To obtain the harmonic components, the time dependence in these equations must be separated out. It was pointed out in Section 37 that Maxwell's equations for the case of a rapidly variable field could be written only for such components. Taking into account that $vk_x = \omega$ and denoting the two-dimensional vector with components k_x, k_y by $\mathbf{q}$, we rewrite ρ and $j = j_x$ in the form

$$\rho = \frac{1}{(2\pi)^3\, v}\int d\tau_q \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t} e^{i(\mathbf{q}\mathbf{r}_\perp + \omega x/v)}$$

$$= \int_{-\infty}^{\infty} \rho\,(\omega)\, e^{-i\omega t}\, d\omega \tag{39.22}$$

$$j_x = \frac{1}{(2\pi)^3} \int d\tau_q \int_{-\infty}^{\infty} d\omega\, e^{-i\omega t} e^{i(\mathbf{q}\mathbf{r}_\perp + \omega x/v)}$$

$$= \int_{-\infty}^{\infty} j_x(\omega)\, e^{-i\omega t} d\omega \qquad (39.23)$$

where the amplitudes $\rho(\omega)$ and $j_x(\omega)$ are apparent from the equations. They are represented as Fourier expansions in terms of $e^{i\mathbf{q}\mathbf{r}_\perp}$.

The obtained charge and current densities should be treated as extraneous with respect to the medium. They should therefore be substituted into the right-hand sides of the wave equations:

$$\nabla^2\varphi(\omega) + \frac{\varepsilon}{c^2}\omega^2\varphi(\omega) = -\frac{4\pi\rho(\omega)}{\varepsilon} \qquad (39.24)$$

$$\nabla^2 A_x(\omega) + \frac{\varepsilon}{c^2}\omega^2 A_x(\omega) = -\frac{4\pi}{c} j_x(\omega) \qquad (39.25)$$

which can be obtained in the conventional way from (38.1) and (38.2) by substituting the potentials for the fields according to [12.34] and [12.35]:

$$\mathbf{H} = \operatorname{curl} \mathbf{A}, \qquad \mathbf{E} = i\omega\mathbf{A} - \operatorname{grad}\varphi \qquad (39.26)$$

Comparing the amplitudes of the Fourier expansions on both sides of Eqs. (39.24) and (39.25), we obtain the Fourier components of the potentials:

$$\varphi(\omega) = \frac{e}{2\pi^2\varepsilon(\omega)v} \int \frac{dk_x\, dk_y \exp[i(\mathbf{q}\mathbf{r}_\perp + \omega x/v)]}{q^2 + \omega^2[v^{-2} - \varepsilon(\omega)/c^2]} \qquad (39.27)$$

$$A_x(\omega) = \frac{e}{2\pi^2 c} \int \frac{dk_k\, dk_y \exp[i(\mathbf{q}\mathbf{r}_\perp + \omega x/v)]}{q^2 + \omega^2[v^{-2} - \varepsilon(\omega)/c^2]} \qquad (39.28)$$

Here we have taken into account that the operator ∇^2 applied to a separate Fourier component with respect to ω and to q multiplies it by $-(q^2 + \omega^2/v^2)$. It is easy to verify that the obtained potentials satisfy the Lorentz condition [12.42], which in the present case has the form

$$-i\frac{\omega\varepsilon}{c}\varphi + \frac{\partial A}{\partial x} = 0 \qquad (39.29)$$

The force acting on a charge is equal to the product of the charge and the electric field taken at its point of location ($\mathbf{r}_\perp = 0$, $x = vt$) with the opposite sign (minus is taken because the field $\mathbf{E}$ is produced by the charge itself). For one Fourier component this constitutes

$$eE(\omega)\,d\omega = -\left(i\omega A(\omega) + i\frac{\omega}{v}\varphi\right) d\omega$$

$$= -\frac{ie^2}{2\pi^2} d\omega \int \frac{dk_x\, dk_y\, \omega(c^{-2} - \varepsilon^{-1}v^{-2})}{q^2 + \omega^2(v^{-2} - \varepsilon/c^2)} \qquad (39.30)$$

This takes into account that the force is parallel to the velocity of the electron. In integrating it is convenient to go over to polar coordinates, replacing $dk_x\,dk_y$ by $2\pi q\,dq$. Furthermore, to the obtained expression for the field we must add a similar term from the expansions (39.20) and (39.21) for the negative frequency ω so as to obtain the whole harmonic component corresponding to $|\omega|$:

$$-eE(\omega)\,d\omega = -\frac{ie^2}{\pi}\,d\omega \int_0^\infty q\,dq\,[f_q(\omega)+f_q(-\omega)] \qquad (39.31)$$

where

$$f_q(\omega)=\frac{\omega\,[c^{-2}-\varepsilon^{-1}(\omega)\,v^{-2}]}{q^2+\omega^2\,[v^{-2}-\varepsilon(\omega)/c^2]}$$

Calculation of the Cerenkov Radiation Intensity. It might appear at first glance that Eq. (39.31) yields zero value for the field. In fact, the correct value of the effect is obtained by means of a limiting process.

We assumed implicitly that the medium was transparent, so that $\varepsilon(\omega)$ has only a real component $\varepsilon'(\omega)$, which, according to (36.14), is an even function of frequency: $\varepsilon'(\omega) = \varepsilon'(-\omega)$. But in that case neither term under the integral in (39.31) has a definite value, since at $v > c\varepsilon^{-1/2}$ the second term in the denominator is negative, and thus there exists a positive quantity $q^2 = \omega^2\,[\varepsilon(\omega)/c^2 - v^{-2}]$ that makes the denominator vanish.

Before calculating the integral, first note that there are no ideally transparent media. The quantity $\varepsilon(\omega)$ always has a small positive imaginary part $\varepsilon''(\omega)$, and $\varepsilon(-\omega)$ correspondingly has an imaginary part $-\varepsilon''(\omega)$, since from (36.15) $\varepsilon''(-\omega) = -\varepsilon''(\omega)$. Thanks to this the denominators at the respective points are not zero but are equal to imaginary quantities of opposite signs. When $\varepsilon''(\omega)$ tends to zero, the integral (39.31) tends to a definite finite value.

Substituting $(1/2)d\xi$ for $q\,dq$ and cancelling in an obvious manner, we calculate the limit at $\lambda \to 0$ of such an expression:

$$\frac{1}{2}\int_0^\infty \frac{d\xi}{\xi-a+i\lambda}-\frac{1}{2}\int_0^\infty \frac{d\xi}{\xi-a-i\lambda}$$

$$=\frac{1}{2}\left(\int_0^a \frac{d\xi}{\xi-a+i\lambda}-\int_a^\infty \frac{d\xi}{a-\xi-i\lambda}-\int_0^a \frac{d\xi}{\xi-a-i\lambda}+\int_a^\infty \frac{d\xi}{a-\xi+i\lambda}\right)$$

We combine the first integral with the third and the second with the fourth to get

$$-i\lambda\int_0^a \frac{d\xi}{(\xi-a)^2+\lambda^2}+i\lambda\int_a^\infty \frac{d\xi}{(a-\xi)^2+\lambda^2}$$

$$=-i\arctan\frac{\xi-a}{\lambda}\Bigg|_0^a-i\arctan\frac{a-\xi}{\lambda}\Bigg|_a^\infty=-2i\arctan\frac{a}{\lambda}$$

As $\lambda\to 0$, we obtain the required limit, which is equal to $-\pi i$. Hence

$$-eE(\omega)\,d\omega=e^2\omega\,d\omega\,[c^{-2}-\varepsilon^{-1}(\omega)\,v^{-2}] \tag{39.32}$$

Note that if for a given frequency $v<c/\varepsilon^{1/2}(\omega)$, the denominators of the integrands do not become zero, and the expression in brackets in the integral (39.31) is equal to zero. The Cerenkov radiation disappears.

The Fourier components in the expansion of the electromagnetic field of a moving charge do not, in general, correspond to any real radiation of electromagnetic waves at an arbitrary angle. The latter always corresponds to the condition $k=\omega\varepsilon^{1/2}/c$. From this it is simple to determine the only angle formed by the direction of the real radiation of given frequency with the velocity vector of the travelling charge. The substitution $\omega=k_xv$, or $w=kv\cos\theta$, was made in the expansions. But if $\mathbf{k}$ is the radiation wave vector, then its magnitude is equal to $\omega\varepsilon^{1/2}/c$, whence

$$\cos\theta=\frac{c}{v\varepsilon^{1/2}} \tag{39.33}$$

From this it is again apparent that radiation can occur only when $v>c/\varepsilon^{1/2}$.

Light Scattering by Fluctuations. No medium is perfectly homogeneous. Thermal motions produce density fluctuations, which implies that a plane electromagnetic wave travelling through a medium is necessarily distorted and scattered by the inhomogeneities.

Let us determine the damping constant of a plane wave due to such scattering in a gaseous medium. Let the number of molecules in a certain volume V have changed by ΔN due to fluctuation. The supplementary dipole moment of the volume V in the field of a plane electromagnetic wave $\mathbf{E}$ due to that density fluctuation is

$$\Delta\mathbf{d}=\mathbf{E}\beta V\,\Delta N \tag{39.34}$$

where β is a coefficient of proportionality between the dipole moment of a molecule and the external field $\mathbf{E}$ (in the assumption that it is

due completely to that field). The quantity β can easily be linked with the dielectric constant of the gas. If the density of the gas is n, the dipole moment of a unit volume (polarization $\mathbf{P}$) in the field $\mathbf{E}$ is $n\beta\mathbf{E}$. Hence the dielectric constant is expressed as follows:

$$\varepsilon(\omega) = 1 + 4\pi n\beta(\omega) \tag{39.35}$$

Since the field is dependent upon time according to a harmonic law, the second derivative of the dipole moment is

$$\Delta\ddot{\mathbf{d}} = -\omega^2\beta(\omega)\,\mathbf{E}\,\Delta N \tag{39.36}$$

The intensity of the scattered radiation is determined from [20.28] as follows:

$$I = \frac{2}{3}\frac{|\Delta\ddot{\mathbf{d}}|^2}{c^3} = \frac{2}{3}\,\omega^4\beta^2\mathbf{E}^2(\Delta N)^2 \tag{39.37}$$

The energy flux of scattered radiation emerging from unit volume, that is, I/V, must be averaged over the fluctuations. As was shown in Exercise 1, Section 10, in a gas $\overline{(\Delta N)^2} = N$.

The damping of the initial plane wave is due not to absorption but to radiation scattering, and is proportional not to the wave amplitude but to its square, that is, to the energy flux. Denoting the damping over the unit length as η, we see that it is equal to the ratio of I/V to the energy flux of the incident radiation $c\mathbf{E}^2/(4\pi)$. The coefficient β is usually replaced by $(\varepsilon - 1)/(4\pi n)$ from (39.35). Hence

$$\eta = \frac{1}{6\pi}\frac{(\varepsilon-1)^2\,\omega^4}{nc^4} \tag{39.38}$$

Thus, fluctuation scattering is the greater the higher the radiation frequency. In the solar spectrum, the atmosphere scatters blue rays most, which explains the blue colour of the sky.

Let us briefly examine the question of polarization of scattered radiation. Equation (39.34) shows that the induced dipole moment is parallel to the field. As is known from [Sec. 20], the electric vector of scattered radiation lies in the same plane as the dipole moment and the direction of scattering. In any case, the vector of the electric field of the incident wave is perpendicular to its direction. Consequently, radiation scattered at right angles to the direction of the incident radiation possesses an electric vector in a plane perpendicular to the incident beam. Given this condition, we see that it is coplanar with the direction of scattering and the induced dipole moment. Therefore natural (nonpolarized) light is scattered in the perpendicular direction, like plane-polarized light. In other directions it is polarized partially, with a preferred direction of the electric vector.

Actually, the quantity β for an individual molecule is a tensor, not a scalar. Its induced dipole moment is not parallel to the field, which is why total polarization does not occur even in a perfectly transparent, unpolluted atmosphere. Besides, one and the same beam may be scattered repeatedly, which also disrupts total polarization.

The human eye does not react to light polarization. But the eye of a bee is sensitive to it. In flight bees orient themselves by the sun, and light from any small unclouded section of the sky is sufficient for them to determine its position in the sky.

PART IV

PHYSICAL KINETICS

40

GENERAL RELATIONSHIPS

No general theory of nonequilibrium states of statistical systems with many degrees of freedom has yet been formulated that can compare favourably with the theory of equilibrium states based on the Gibbs distribution (Sec. 7). There are, however, relationships that hold for states approaching equilibrium. In deducing them it is usually assumed that a system is subject to some external action that disturbs its equilibrium. If the action is not strong, deviation from equilibrium can in most cases be described with the help of linear expansions with respect to the amplitude of the disturbance. In that case there are found to exist common relationships between the coefficients of the linear forms expressing the rate of processes occurring in the system in terms of the amplitudes of the disturbances applied to it.

The Onsager Reciprocity Theorem. Suppose that certain static external actions disturb a system's statistical equilibrium. If, for example, there is an electric potential gradient in a conductor or a temperature gradient in a medium with arbitrary properties, equilibrium is impossible until the field or temperature gradient becomes zero.

As pointed out in Section 31, these gradients may disturb the equilibrium in two ways: a potential gradient produces not only electric current but heat flux as well; temperature gradients also cause fluxes of both types.

In an anisotropic conductor a potential gradient along one coordinate axis may induce current along the other axes, provided the conductivity tensor σ_{ik} has a nonzero off-diagonal elements.

L. Onsager showed that symmetry relations deriving from very general properties of statistical systems exist between the off-diagonal components of σ_{ik}, or the "coupling coefficients", expressing the heat flow in terms of the potential gradient, and the electric current in terms of the temperature gradient. Let us introduce the concept of the mean value of a certain additive, but spatially nonuniform, quantity x in volume V:

$$x = \frac{1}{V} \int x(\mathbf{r})\, dV \tag{40.1}$$

where $x(\mathbf{r})$ is the local value of this quantity at point $\mathbf{r}$ of volume V. Then $\dot{x} = dx/dt$ characterizes the flux of the given additive quantity across the surface encompassing the volume. For example, if x is an electric charge, then $\dot{x}$ is current; if x is energy, $\dot{x}$ is total heat flux, etc. The x quantities may, without restricting the generalization, be counted off from their median, that is, equilibrium, values, assuming $\bar{x} = 0$.

In thermodynamic equilibrium the entropy S of a given volume is maximum, so that the equilibrium value of the quantity x_k is found from the condition

$$X_k \equiv \frac{\partial S}{\partial x_k} = 0 \tag{40.2}$$

If $X_k \neq 0$, equilibrium does not set in. The quantities X_k are the "forces" disturbing the system's equilibrium. When the nonequilibrium is weak, the quantities $\dot{x}_i$, which vanish at equilibrium, are linear functions of X_k. Using the summation convention, we write the equation linking $\dot{x}_i$ with X_k:

$$\dot{x}_i = \alpha_{ik} X_k \tag{40.3}$$

This equation defines the coefficients α_{ik}, usually called the *phenomenological coefficients*. The *Onsager reciprocity theorem* asserts that $\alpha_{ik} = \alpha_{ki}$.

To prove this, consider two mean quantities:

$$\overline{x_i(t)\, x_k(t+\tau)} \qquad \text{and} \qquad \overline{x_i(t+\tau)\, x_k(t)}$$

In the first, the quantity x_k is taken at the later instant in averaging, in the second, the quantity x_i. If there is no external magnetic field applied to the system, the properties of the system are symmetrical with respect to the operation $t \to -t$, that is, with respect to time inversion. In an external field we must, in addition to t, change the sign of $\mathbf{H}$.

In a system symmetrical with respect to time inversion $t \to -t$ it is immaterial whether x_i or x_k is taken at the later instant. Hence the two mean values are equal:

$$\overline{x_i(t)\, x_k(t+\tau)} = \overline{x_i(t+\tau)\, x_k(t)} \tag{40.4a}$$

Assuming the time interval τ small, we expand x_k and x_i into a series in powers of τ and restrict ourselves to the first term of the expansion. Then Eq. (40.4a) should be written as follows:

$$\overline{x_i \dot{x}_k} = \overline{\dot{x}_i x_k} \tag{40.4b}$$

To perform the averaging, we recall that $\exp[S(x_1, x_2, \ldots, x_n)]$ is the probability of the state with the given values of the quantity x_i. Therefore in more expanded form Eq. (40.4b) appears as

$$\int x_i \dot{x}_k e^S \prod_l dx_l = \int \dot{x}_i x_k e^S \prod_l dx_l \tag{40.5}$$

We substitute into this the expansions $\dot{x}_i$, $\dot{x}_k$ (40.3), taking into account definition (40.2), and get

$$\int x_i \alpha_{kj} \frac{\partial S}{\partial x_j} \prod_l dx_l = \int x_k \alpha_{ij} \frac{\partial S}{\partial x_j} \prod_l dx_l \tag{40.6}$$

Consider the integral

$$I_{ij} \equiv \int x_i \frac{\partial S}{\partial x_j} e^S \prod_l dx_l \tag{40.7}$$

If $i \neq j$, the integral with respect to x_i, x_j yields

$$I_{ij} = \int x_i \prod_{l \neq j} dx_l \int \frac{\partial S}{\partial x_j} e^S dx_j$$

$$= \int x_i \prod_{l \neq j} dx_l \, e^S \Big|_{x_j=-\infty}^{x_j=\infty} = 0$$

This follows from the infinitesimal probability of an infinitely large deviation of the quantities from their equilibrium values. If $i = j$, we obtain

$$\int \prod_{l \neq i} dx_l \int x_i \frac{\partial S}{\partial x_i} dx_i = \int \prod_{l \neq i} dx_l \int x_l \frac{\partial e^S}{\partial x_i} dx_i$$

$$= \int \prod_{l \neq i} dx_l \left(e^S \Big|_{x_i=-\infty}^{x_i=\infty} - \int e^S \prod_l dx_l \right) = -1$$

The expression integrated over x_i becomes zero at the limits, and the remaining integral is equal to -1 by virtue of the normalization of probabilities. Hence

$$\int x_i X_j e^S \prod_l dx_l = \int x_i \frac{\partial e^S}{\partial x_j} \prod_l dx_l = -\delta_{ij} \tag{40.8}$$

Substituting this into (40.6), we obtain $\alpha_{kj}\delta_{ij} = \alpha_{ij}\delta_{kj}$, or

$$\alpha_{ij} = \alpha_{ji} \tag{40.9}$$

as was initially asserted. This equation expresses the symmetry of phenomenological coefficients. In developing it explicit use was made of the reversibility of spontaneous fluctuations with respect to time inversion, which was mentioned in Section 10.

Application of the Onsager Reciprocity Theorem to Thermoelectric Phenomena. In Section 31 we wrote the expressions (31.26) and (31.30) for current and heat flow, which we shall rewrite as applicable to the flux densities $\mathbf{j}$ and $\mathbf{q}$:

$$\mathbf{j} = \sigma(\mathbf{E} - \alpha \operatorname{grad} \theta), \qquad \mathbf{q} - \varphi\mathbf{j} = \beta\mathbf{E} - \gamma \operatorname{grad} \theta$$

The meaning of the notation is made apparent by a comparison with Section 31.

The energy evolved in a unit volume in unit time is

$$\operatorname{div} \mathbf{q} = \operatorname{div}(\mathbf{q} - \varphi\mathbf{j}) + \operatorname{div} \varphi\mathbf{j} \tag{40.10}$$

But since $\operatorname{div} \mathbf{j} = 0$ and $-\operatorname{grad} \varphi = \mathbf{E}$,

$$\operatorname{div} \mathbf{q} = \operatorname{div}(\mathbf{q} - \varphi\mathbf{j}) - (\mathbf{E}\cdot\mathbf{j})$$

The divergence of the energy flux is equal to the energy change per unit time in unit volume taken with the minus sign: $\operatorname{div} \mathbf{q} = -\partial Q/\partial t$. The state of the conductor does not change in the evolution of energy, that is, the external parameters are constant. Thus, the energy represents heat according to its general definition in Section 8. But then the change in entropy per unit time in the volume of the conductor is, from the relationship $dS = dQ/\theta$, equal to

$$\frac{dS}{dt} = -\int \frac{\operatorname{div} \mathbf{q}}{\theta} dV = \int \frac{\mathbf{Ej}}{\theta} dV - \int \frac{\operatorname{div}(\mathbf{q} - \varphi\mathbf{j})}{\theta} dV \tag{40.11}$$

The second integral on the right transforms by parts to the following form:

$$-\int \frac{\operatorname{div}(\mathbf{q} - \varphi\mathbf{j})}{\theta} dV = \int (\mathbf{q} - \varphi\mathbf{j}) \frac{\operatorname{grad} \theta}{\theta^2} dV \tag{40.12}$$

whence we obtain

$$\frac{dS}{dt} = \int \frac{\mathbf{Ej}}{\theta} dV + \int \frac{\mathbf{q} - \varphi\mathbf{j}}{\theta^2} \operatorname{grad} \theta \, dV \tag{40.13}$$

Now compare this expression with the general definition (40.2) of the parameters characterizing a nonequilibrium state.

In equilibrium the current **j** and heat flux $\mathbf{q} - \varphi\mathbf{j}$ are zero. Deviation from equilibrium occurs at the expense of **E** and grad θ. The time derivatives of the parameters x_i, that is $\dot{x}_i$, are, in the present case, **j** and $\mathbf{q} - \varphi\mathbf{j}$. For example, a steady current across a unit surface is equal to the charge of current carriers in unit volume, ρ, multiplied by the displacement velocity of the "electric centre" $\dot{\mathbf{a}}$ of those charges [16.22]. Taking also into account that ρ = constant, we obtain $\mathbf{j} = \rho(\partial\mathbf{a}/\partial t) = (\partial/\partial t)\,\rho\mathbf{a}$. The heat flux can be determined similarly.

Now, taking into account that

$$\frac{dS}{dt} = \frac{\partial S}{\partial x_i}\frac{\partial x_i}{\partial t} \equiv X_i\dot{x}_i \tag{40.14}$$

we see that the quantities X_i, referred in the present case to unit volume, are

$$X_{\mathbf{E}} = -\frac{\mathbf{E}}{\theta}, \qquad X_\theta = \frac{\operatorname{grad}\theta}{\theta^2} \tag{40.15}$$

We rewrite the initial transport equations as follows:

$$\mathbf{j} = \sigma\mathbf{E} - \sigma\alpha\theta^2\frac{\operatorname{grad}\theta}{\theta^2} = -\sigma X_{\mathbf{E}} - \sigma\alpha\theta^2 X_\theta \tag{40.16a}$$

$$\mathbf{q} - \varphi\mathbf{j} = \beta\theta\frac{\mathbf{E}}{\theta} - \gamma\operatorname{grad}\theta = -\beta\theta X_{\mathbf{E}} - \gamma X_\theta \tag{40.16b}$$

Applying the Onsager reciprocity theorem (40.9), we obtain the equation for the coupling coefficients:

$$\beta = \sigma\alpha\theta \tag{40.17}$$

Equation (31.30), analogous to this one, was proved with the help of the second law of thermodynamics, which was an essentially weak point in developing thermoelectric relationships. By contrast, Eq. (40.17) was obtained quite strictly.

Resolution of Fluctuations into Harmonic Components. Most general relationships can also be obtained in the kinetics of rapidly variable processes if they weakly affect the equilibrium of the system in which they take place. Physical quantities deviate from their equilibrium values rapidly but with small amplitudes. As was shown in Section 36, in describing such processes it is convenient to resolve the variables into components varying according to a harmonic law.

We shall develop one important relationship relating to the harmonic components of quantities.

Let a system involve a certain quantity $a(t)$ varying spontaneously according to a random law, for example, as a consequence of heat fluctuations. In particular, this may be the readings of a galvanometer switched into a closed circuit in the absence of external emf. Heat fluctuations may produce an irregular emf and current of any sign in a circuit. The temperature and concentration of current carriers fluctuate at different sections of the circuit, which causes a variable emf.

Let us represent an irregularly varying quantity $a(t)$ in the form of a Fourier integral:

$$a(t) = \int_{-\infty}^{\infty} a(\omega) e^{i\omega t} d\omega \tag{40.18}$$

Here, $a(\omega)$ are called the *Fourier components* of $a(t)$. Since $a(t)$ does not vanish at $t = \pm\infty$, the question may arise of whether expansion (40.18) is legitimate. It should be interpreted in the following sense. Let the current in the circuit be resolved into harmonic components with the help of a wave analyzer sensitive to frequencies in the interval $\Delta\omega$. Then $a(\omega)$ is the reading at frequency ω within the interval $\Delta\omega$. If the integral $\Delta\omega$ is stated, this is a definite quantity.

Since $a(t)$ is a real quantity,

$$a^*(t) = a(t) = \int_{-\infty}^{\infty} a^*(\omega) e^{-i\omega t} d\omega = \int_{-\infty}^{\infty} a(\omega) e^{i\omega t} d\omega$$

Comparing the amplitudes at $e^{-i\omega t}$, we obtain

$$a^*(\omega) = a(-\omega) \tag{40.19}$$

Let us now show how the mean square of $a(t)$ is expressed in terms of its Fourier components. By definition

$$\overline{[a(t)]^2} = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} a^2(t)\, dt \tag{40.20}$$

Substituting the expansion (40.18) and interchanging the order of integration over frequency and time, we obtain

$$\overline{[a(t)]^2} = \int_{-\infty}^{\infty} a(\omega)\, d\omega \int_{-\infty}^{\infty} a(\omega')\, d\omega' \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e^{i(\omega+\omega')t}\, dt$$

If $\omega + \omega' \neq 0$, the integral of $e^{i(\omega+\omega')t}$ within infinite limits tends to zero. But if $\omega + \omega' = 0$, the integral increases as $2T$.

Consequently

$$\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}e^{i(\omega+\omega')t}\,dt=1$$

Taking this into account and, in addition, substituting $\omega'=-\omega$, we obtain

$$\begin{aligned}\overline{a^2}&=\int d\omega'\int_{-\infty}^{\infty}a(\omega)\,a(-\omega)\,d\omega\\&=\int d\omega'\int_{-\infty}^{\infty}|a(\omega)|^2\,d\omega\end{aligned}\tag{40.21}$$

Here we made use of condition (40.19). The integral over ω does not depend on ω'. In the integral over ω' the limits have not been substituted. This means that the mean value $\overline{a^2}$ is proportional to the frequency interval $\Delta\omega$ in the Fourier expansion, which represents the fluctuating quantity. The interval $\Delta\omega$ has already been defined as the *transmission band* of the wave analyzer measuring $a(\omega)$. Hence

$$\overline{a^2}=\Delta\omega\int_{-\infty}^{\infty}|a(\omega)|^2\,d\omega\tag{40.22}$$

The explicit dependence on $\Delta\omega$ is removed if we put $a'(\omega)=(\Delta\omega)^{1/2}a(\omega)$. Thus $\Delta\omega$ is involved only in the normalization of the Fourier components, while the integral in the expansion extends over an infinite frequency interval.

Voltage Fluctuations in Linear Electric Circuits. Let us now apply (40.22) to a linear electric circuit with elements, which have capacitance, self-inductance (and, in the general case, mutual inductance as well), and resistance. Suppose that the capacitances, as well as the self-inductances and mutual inductances, depend only on the geometry and configuration of the conductors and are therefore not subject to statistical fluctuations. For this the capacitors should not include dielectrics and the inductance coils should not include ferromagnetic cores. The resistor on which dissipation of the energy of the electromagnetic field occurs, that is, transforms into heat, is essentially a statistical element in the circuit. Generalization for the case when dissipation occurs in all elements of the circuit presents no difficulty.

If the resistor is the statistical element in the circuit, the fluctuations occur in it. In particular, in the absence of an external emf a continuously varying random potential difference appears spon-

taneously on the ends of the section of the circuit containing the resistor. This phenomenon is conventionally called *noise*: given appropriate amplification this is just what is heard in an earphone introduced into the circuit. The interval $\Delta\omega$ represents the transmission width of the amplifier.

H. Nyquist showed in 1928 that noise level is directly proportional to resistance (in the most general case, to the real part of the imped-

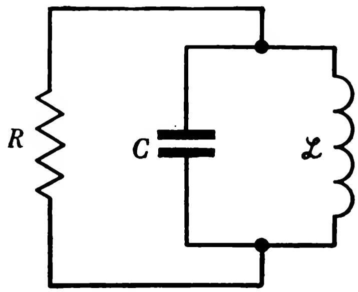


Figure 52

ance). Thus a kinetic characteristic of a system (resistance) was for the first time linked with the fluctuations taking place in that system in the equilibrium state.

To determine the relationship, consider a circuit of the type shown in Figure 52. It is a resonant circuit composed of a capacitance element C and an inductance element $\mathscr{L}$ in series with a resistor R.

We shall first determine the random emf in the circuit, if the Fourier component of the voltage fluctuation at the ends of the resistor R is $V_R(\omega)$. Like $a'(\omega)$, the Fourier components are defined together with the square root of the frequency interval $\Delta\omega$. Consequently

$$\overline{\mathscr{E}^2} = \int_{-\infty}^{\infty} |\mathscr{E}(\omega)|^2 \, d\omega \tag{40.23}$$

where $\mathscr{E}(\omega)$ denotes the Fourier component of the emf in the contour multiplied by $(\Delta\omega)^{1/2}$.

Let us express $\mathscr{E}(\omega)$ in terms of $V_R(\omega)$. As was shown in Section 35, linear *a-c* circuits are calculated in the same way as *d-c* circuits. Denoting inductive and capacitive reactance by the symbols R_1 and R_2, we obtain the resistance of the section of the circuit in which they are joined in parallel:

$$R' = (R_1^{-1} + R_2^{-1})^{-1}$$

From Ohm's law it is apparent that the ratio of the emf in the whole circuit to the voltage drop across the resistor R is

$$\frac{\mathscr{E}(\omega)}{V_R(\omega)}=\frac{R'}{R+R'}=\frac{1}{1+R/R'}=\frac{1}{1+R/R_1+R/R_2}$$

Substituting $R_1 = -i\omega\mathscr{L}$, $R_2 = i(\omega C)^{-1}$ (Sec. 35), we obtain the square of the modulus of the Fourier component of the emf:

$$|\mathscr{E}(\omega)|^2=\frac{|V_R(\omega)|^2}{1+R^2[(\omega C)^{-1}-\omega\mathscr{L}]^2} \tag{40.24}$$

We express the quantities involved in this formula in terms of the circuit's resonance frequency $\omega_0 = (LC)^{-1/2}$:

$$|\mathscr{E}(\omega)|^2=\frac{|V_R(\omega)|^2}{1+R^2C^2\omega_0^2(\omega/\omega_0-\omega_0/\omega)^2} \tag{40.25}$$

We now make use of the relationship (40.22). We rewrite it as follows (remembering that $(\Delta\omega)^{1/2}$ is involved in the definition of $\mathscr{E}(\omega)$ and $V_R(\omega)$):

$$C\overline{\mathscr{E}^2}=C\int_{-\infty}^{\infty}|\mathscr{E}(\omega)|^2\,d\omega \tag{40.26}$$

It was shown in Section 35 that an oscillator circuit is equivalent to a linear harmonic oscillator. Its mean energy value in thermal equilibrium represents the double value of the mean potential energy. Making use now of (3.5), we find that

$$C\overline{\mathscr{E}^2}=h\omega_0\left(\frac{1}{2}+\frac{1}{e^{h\omega_0/\theta}-1}\right)=\frac{h\omega_0}{2}\coth\frac{h\omega_0}{2\theta} \tag{40.27}$$

Assuming the resonance at $\omega = \omega_0$ to be sufficiently sharp (this can always be achieved by a judicious choice of the inductive and capacitive reactances), we can take the square of the amplitude outside the integral sign at $|\mathscr{E}(\omega)|^2 = |\mathscr{E}(\omega_0)|^2$. Integrating[1], we

[1] The integral is computed by means of the following substitution:

$$\frac{\omega}{\omega_0}-\frac{\omega_0}{\omega}=z, \qquad dz=\frac{\omega_0}{2}\left(1\pm\frac{z}{(z^2+4)^{1/2}}\right)$$

The upper sign corresponds to the frequency interval $0 \leqslant \omega \leqslant \infty$, and the lower to the frequency interval $-\infty \leqslant \omega < 0$. When we add together the integrals referring to both frequency intervals, the root cancels out and there remains

$$\int_{-\infty}^{\infty}\frac{\omega_0\,dz}{1+R^2C^2\omega_0^2z^2}=\frac{\pi}{RC}$$

obtain

$$C\,|\mathscr{E}(\omega_0)|^2 \int\limits_{-\infty}^{\infty} \frac{d\omega}{1+R^2C^2\omega_0^2\,(\omega/\omega_0-\omega_0/\omega)^2} = \frac{\pi}{R}\,|\mathscr{E}(\omega_0)|^2 \quad (40.28)$$

Equating this result to expression (40.27), we arrive at the required relationship:

$$|\mathscr{E}(\omega_0)|^2 = R\frac{h\omega_0}{2\pi}\coth\frac{h\omega_0}{2\theta} \quad (40.29)$$

At sufficiently high temperature, when $h\omega_0/(2\theta) \ll 1$,

$$\frac{h\omega_0}{2\theta}\coth\frac{h\omega_0}{2\theta} \approx 1 \quad (40.30)$$

and the obtained formula reduces to the form initially deduced by Nyquist:

$$|\mathscr{E}(\omega_0)|^2 = \frac{R\theta}{\pi} \quad (40.31)$$

Electric Conductivity in Rapidly Variable Fields. Nyquist's relationship (40.31) is generalized for rapidly variable fields with periods comparable with the settling time of the current in the circuit. In these conditions electric conductivity exhibits dispersion. In the most general case the linear relationship between field and current is of an integral type similar to (36.1):

$$j_\alpha(t) = \int\limits_0^\infty \sigma_{\alpha\beta}(\tau)\,E_\beta(t-\tau)\,d\tau \quad (40.32)$$

where α and β are tensor indices. The component $E_\alpha(t)$ is not involved here on the right-hand side, since in (36.1) it contributed to the equality of **E** and **D** in vacuum.

If $E_\beta(t)$ depends upon time according to the harmonic law

$$E_\beta(t) = E_{0\beta}(\omega)\,e^{-i\omega t}$$

the relationship between field and current becomes a simple proportionality:

$$j_\alpha(\omega) = \left(\int\limits_0^\infty \sigma_{\alpha\beta}(\tau)\,e^{i\omega t}\,d\tau\right) E_\beta(\omega) \quad (40.33)$$

Hence, the integral in parentheses represents the expression for electric conductivity at the frequency ω.

R. Kubo expressed $\sigma_{\alpha\beta}(\tau)$ in terms of the averaged correlations of the current in a conductor in statistical equilibrium for various instants of time (for the definition of correlation between quantities

see Section 10):

$$\sigma_{\alpha\beta}(\tau) = \frac{1}{\theta} \int \langle j_\beta(\mathbf{r}, t)\, j_\alpha(0, 0)\rangle\, dV \tag{40.34}$$

Here the angle brackets denote a mean quantity; the symbol has been adopted instead of the conventional overhead bar because the averaging is carried out by quantum mechanical methods, as in [Sec. 25]. The quantity $j_\alpha(0, 0)$ denotes the random equilibrium fluctuation of the current in the x_α direction at the origin of the coordinate system and at the initial time. In a homogeneous conductor not subject to external effects both origins are chosen arbitrarily. The quantity $j_\beta(\mathbf{r}, t)$ is a like fluctuation of the current at point $\mathbf{r}$ and at time t in the x_β direction. Obviously $\langle j_\beta(\mathbf{r}, t)\rangle = \langle j_\alpha(0, 0)\rangle = 0$. Since fluctuations do not occur instantaneously but over a time interval necessary for relaxation in the system, at not too great values of τ, $j_\beta(\mathbf{r}, \tau)$ and $j_\alpha(0, 0)$ are not independent variables, and the mean of their product is not zero. At $\alpha \neq \beta$ it is obviously nonzero only in a conductor with a specific anisotropy. This mean determines the conductivity tensor in the integral relationship (40.32), while the Fourier component of it yields conductivity at the corresponding frequency.

The difference between (40.34) and Nyquist's relationship consists primarily in that the latter involves mean quadratic fluctuations for the same instant, while the *Kubo formula* refers to the mean of the fluctuations occurring at different times. That is why it does not involve the frequency interval $\Delta\omega$.

Some Necessary Formulas. To develop the Kubo formula we will have to recall some quantum mechanical formulas, develop others, and bring them together.

We shall begin with the *density matrix* [27.29]:

$$\rho(x', x) = \sum_n w_n \psi_n^*(x')\, \psi_n(x) \tag{40.35}$$

where w_n is the probability of the system occurring in the nth state, and $\psi_n(x)$ is the wave function of that state.

A description using the density matrix is used with respect to open systems which, if isolated from external actions, may occur in pure states $\psi_n(x)$. In particular, a system in statistical equilibrium with the surroundings is described by the density matrix

$$\rho_0 = e^{(F - \hat{\mathscr{H}})/\theta} \tag{40.36}$$

(see Exercise 2, Section 7). Here $\hat{\mathscr{H}}$ is the Hamiltonian of the system.

The density matrix satisfies equation [27.48]:

$$\frac{d\rho}{dt} = \frac{i}{h}(\hat{\mathscr{H}}^*\rho - \hat{\mathscr{H}}\rho)$$

which is conveniently rewritten in somewhat different form, taking advantage of the fact that $\hat{\mathscr{H}}$ is a Hermitian operator. Namely, $\hat{\mathscr{H}}^* = \tilde{\hat{\mathscr{H}}}$, where the tilde denotes a transposition, that is, when the rows and columns of the corresponding matrix are interchanged. But by the definition of a transposed operator [37.22], $\tilde{\hat{\mathscr{H}}}\rho = \rho\hat{\mathscr{H}}$. It follows then that

$$\frac{d\rho}{dt} = \frac{i}{h}(\rho\hat{\mathscr{H}} - \hat{\mathscr{H}}\rho) \tag{40.37}$$

Here the definition of the derivative with respect to time differs from the derivative of a certain operator $\hat{\lambda}$ with respect to time only in sign.

The density matrix is used to compute the mean values of quantities in states described in terms of definition (40.35) rather than by wave functions. The method of computing mean values is conveniently applied to open systems, notably those in statistical equilibrium. That, actually, is why the matrix $\rho(x', x)$ is introduced. From [27.36], the mean value of the quantity λ, to which the matrix $\lambda_{xx'}$ corresponds in the x-representation, is

$$\begin{aligned}\langle\lambda\rangle &= \int dx \int dx' \, \lambda_{xx'}\rho(x', x) \\ &= \int dx \, (\lambda\rho)_{xx} = \mathrm{Tr}\,(\lambda\rho) = \mathrm{Tr}\,(\rho\lambda)\end{aligned}$$

where the symbol Tr represents the diagonal sum [27.34] of the matrix.

We shall now obtain two more general formulas for the operator $\hat{\lambda}(t)$. First we find in integrated form the equation that gives the time dependence of $\hat{\lambda}$, that is, we express $\hat{\lambda}(t)$ from the differential equation

$$\frac{d\hat{\lambda}}{dt} = \frac{i}{h}(\hat{\mathscr{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathscr{H}})$$

At time t' the operator $\hat{\lambda}$ has the form

$$\hat{\lambda}(t') = e^{i\hat{\mathscr{H}}(t'-t)/h}\hat{\lambda}(t)\, e^{-i\hat{\mathscr{H}}(t'-t)/h} \tag{40.38}$$

To make sure of this it is sufficient to differentiate (40.38) with respect to time, taking into account that $\hat{\mathscr{H}}$ does not commute with $\hat{\lambda}$. Therefore the derivative of the second exponential function with respect to time is written to the right of $\hat{\lambda}$. Thus

$$\frac{d\hat{\lambda}}{dt'} = \frac{i}{h} e^{i\hat{\mathscr{H}}(t'-t)/h}(\hat{\mathscr{H}}\hat{\lambda} - \hat{\lambda}\hat{\mathscr{H}})\, \hat{e}^{-i\hat{\mathscr{H}}(t'-t)/h}$$

Assuming that $t' = t$, we come to the initial equation.

The second equation we shall need makes it possible to express the commutator of a certain operator $\hat{V}$ with the operator $e^{-\beta\hat{\mathscr{H}}}$ in terms of a commutator of $\hat{V}$ with $\hat{\mathscr{H}}$. This is an identity of general form, in which the specific properties of the Hamiltonian are not used. Namely

$$\hat{V}e^{-\beta\hat{\mathscr{H}}} - e^{-\beta\hat{\mathscr{H}}}\hat{V} = e^{-\beta\hat{\mathscr{H}}}\int_0^\beta d\beta' e^{\beta'\hat{\mathscr{H}}}(\hat{V}\hat{\mathscr{H}} - \hat{\mathscr{H}}\hat{V})\, e^{-\beta'\hat{\mathscr{H}}} \tag{40.39}$$

We differentiate both sides of the equation with respect to β to get

$$\hat{V}e^{-\beta\hat{\mathscr{H}}}\hat{\mathscr{H}} + e^{-\beta\hat{\mathscr{H}}}\hat{\mathscr{H}}\hat{V}$$
$$= e^{-\beta\hat{\mathscr{H}}}\hat{\mathscr{H}}\int_0^\beta d\beta' e^{\beta'\hat{\mathscr{H}}}(\hat{V}\hat{\mathscr{H}} - \hat{\mathscr{H}}\hat{V})e^{-\beta'\hat{\mathscr{H}}} - (\hat{V}\hat{\mathscr{H}} - \hat{\mathscr{H}}\hat{V})e^{-\beta\hat{\mathscr{H}}}$$

Substituting in place of the integral its expression from (40.39), we find that both obtained expressions are the same. Consequently, (40.39) is valid at $\beta = 0$, and also yields an identity after differentiating with respect to β. Therefore it is satisfied at all β's.

The Kubo Formula. Let a weak, time dependent disturbance $\hat{V}(x, t)$ be applied to a system with the Hamiltonian $\hat{\mathscr{H}}$ in statistical equilibrium. Then Eq. (40.37) for the density matrix has the form

$$\frac{d\rho}{dt} = \frac{i}{h}(\rho\hat{\mathscr{H}} - \hat{\mathscr{H}}\rho) + \frac{i}{h}(\rho\hat{V} - \hat{V}\rho)$$

Assuming the disturbance to be weak, we resolve ρ into two parts:

$$\rho = \rho_0 + \rho_1$$

where ρ_0 is defined by formula (40.36), and ρ_1 is a small increment linear in $\hat{V}$. Obviously, if $\hat{V} = 0$, then $d\rho_0/dt = 0$ since ρ_0 commutes with $\hat{\mathscr{H}}$. That, of course, is how it should be in equilibrium. Neglecting the term quadratic in $\hat{V}$, we obtain

$$\frac{d\rho_1}{dt} = \frac{i}{h}(\rho_1\hat{\mathscr{H}} - \hat{\mathscr{H}}\rho_1) + \frac{i}{h}(\rho_0\hat{V} - \hat{V}\rho_0) \tag{40.40}$$

This nonhomogeneous linear equation has the following solution:

$$\rho_1(t) = -\frac{i}{h}\int_{-\infty}^{t} dt'\, e^{i\hat{\mathscr{H}}(t'-t)/h}(\hat{V}\rho_0 - \rho_0\hat{V})\, e^{-i\hat{\mathscr{H}}(t'-t)/h} \tag{40.41}$$

By differentiating (40.41) with respect to time it is easy to see that Eq. (40.40) is satisfied: the derivative with respect to the upper

integration limit yields the term $i(\rho_0\hat{V} - \hat{V}\rho_0)/h$, and the derivative with respect to t under the integral, the term $i(\rho_1\hat{\mathcal{H}} - \hat{\mathcal{H}}\rho_1)/h$, as in the proof of (40.38).

Now let us replace the commutator $(\hat{V}\rho_0 - \rho_0\hat{V})$ using identity (40.39), in which we put $\beta = 1/\theta$. The constant factor of ρ_0, that is $e^{F/\theta}$, is irrelevant. By definition, $\hat{V}(x, t)$ depends only on the coordinates of the system and not on its momenta, so that the operator $\hat{V}$ does not commute with the kinetic energy operator only.

From the quantum mechanical equations of motion [27.8] it follows that the commutator of $\hat{V}$ with the kinetic energy operator represents the change in kinetic energy per unit time in the field $\hat{V} = \sum e\varphi_i$, that is, the work done by the external field $\mathbf{E} = -\operatorname{grad}\varphi$ on the system; $\varphi_i = \varphi(\mathbf{r}_i)$. We determine this from the properties of commutators. The kinetic energy operator for all electrons of the system is $\sum \hat{p}_i^2/(2m)$. The commutator of $\sum \hat{p}_i^2/(2m)$ with $\hat{V}$ is

$$\sum_i \frac{1}{2m}(\hat{p}_i^2\hat{V} - \hat{V}\hat{p}_i^2) = \sum_i \frac{1}{2m}(\hat{p}_i^2\hat{V} - \hat{\mathbf{p}}_i\hat{V}\hat{\mathbf{p}}_i + \hat{\mathbf{p}}_i\hat{V}\hat{\mathbf{p}}_i - \hat{V}\hat{\mathbf{p}}_i^2)$$

$$= \frac{h}{2mi}\sum_i \left(\hat{\mathbf{p}}_i\frac{\partial\hat{V}}{\partial\mathbf{r}_i} - \frac{\partial\hat{V}}{\partial\mathbf{r}_i}\hat{\mathbf{p}}_i\right)$$

(see Exercise 2, [Sec. 24]). Assuming the external electric field to be spatially uniform, $\partial\hat{V}/\partial\mathbf{r}_i = -e\mathbf{E}$ (where $\mathbf{E}$ is no longer an operator but a constant vector). The expression $\hat{p}_i/m$ is the velocity operator of an electron. In the last commutator we go over from the sum to the integral over the volume of the conductor:

$$\sum_i e\frac{\hat{\mathbf{p}}_i}{m}\mathbf{E} = \mathbf{E}\int \sum_i \frac{e\hat{\mathbf{p}}_i}{m}\delta(\mathbf{r}_i - \mathbf{r})\,dV$$

It is now apparent that the integrand is the operator of the charge density multiplied by their velocity operator, that is, it represents the operator of the current density $\hat{\mathbf{j}}(\mathbf{r})$. We finally obtain

$$\frac{d}{dt}\sum_i \frac{\hat{p}_i^2}{2m} = \frac{i}{h}(\hat{\mathcal{H}}\hat{V} - \hat{V}\hat{\mathcal{H}}) = \mathbf{E}\int \hat{\mathbf{j}}(\mathbf{r}, t)\,dV$$

In accordance with the correspondence principle it follows that the change in kinetic energy per unit time is equal to the work done on the system, the equation being understood in the operator sense.

Substituting the obtained formula into the right-hand side of (40.39), we obtain

$$\hat{V}\rho_0 - \rho_0\hat{V} = -\frac{h}{i}\mathbf{E}e^{\beta\hat{\mathcal{H}}}\int_0^{\beta} d\beta' e^{\beta'\hat{\mathcal{H}}}\int \hat{\mathbf{j}}\, dV\, e^{-\beta'\hat{\mathcal{H}}} \tag{40.42}$$

After this, from formula (40.41) we find the required correction to the density matrix, ρ_1 (owing to the external field $\mathbf{E}$), which is

$$\int_{-\infty}^{t} dt'\, \mathbf{E}(t')\left[e^{i\hat{\mathcal{H}}(t'-t)/h}\rho_0\int_0^{\beta} d\beta'\, e^{\beta'\hat{\mathcal{H}}}\right.$$
$$\left.\times\int dV\, \hat{\mathbf{j}}e^{-\beta'\hat{\mathcal{H}}-i\hat{\mathcal{H}}(t'-t)/h}\right] \tag{40.43}$$

Further simplification can be carried out assuming the temperature to be sufficiently high in comparison with all quantities of the type h/τ_n, where τ_n is one of the characteristic relaxation times of the system. If we make this assumption, then in the operator

$$e^{i\hat{\mathcal{H}}(t'-t)/h+\beta'\hat{\mathcal{H}}}\hat{\mathbf{j}}e^{-i\hat{\mathcal{H}}(t'-t)/h-\beta'\hat{\mathcal{H}}}$$

involved in (40.43) we can legitimately neglect the real part of the exponents $\beta'\hat{\mathcal{H}} \equiv \hat{\mathcal{H}}/\theta$ in comparison with the imaginary part $i\hat{\mathcal{H}}\,(t'-t)/h$, since $t'-t$ is effectively of the order of magnitude of τ_n. But in that case we must apply formula (40.38), so that

$$e^{i\hat{\mathcal{H}}(t'-t)/h}\hat{\mathbf{j}}(t)\, e^{-i\hat{\mathcal{H}}(t'-t)/h} = \hat{\mathbf{j}}(t') \tag{40.44}$$

and integration with respect to β' yields simply $\beta = 1/\theta$.

As a result the correction to the density matrix turns out to be

$$\rho_1 = \frac{\rho_0}{\theta}\int_{-\infty}^{t} dt'\, \mathbf{E}(t')\int \hat{\mathbf{j}}(\mathbf{r},\, t')\, dV \tag{40.45}$$

Let us use this formula and (40.37) to calculate the mean current at point $\mathbf{r} = 0$ at time t. It is obvious that the undisturbed matrix ρ_0 makes no contribution, so that

$$\begin{aligned}\langle \hat{j}_\alpha(t)\rangle &= \mathrm{Tr}\,(\rho_1\hat{j}_\alpha)\\ &= \frac{1}{\theta}\int_{-\infty}^{t} dt'\, E_\beta(t')\int dV\, \mathrm{Tr}\,[\rho_0\hat{j}_\beta(\mathbf{r},\, t')\,\hat{j}_\alpha(0,\, t)]\\ &= \frac{1}{\theta}\int_{-\infty}^{t} dt'\, E_\beta(t')\int dV\, \langle j_\beta(\mathbf{r},\, t')\, j_\alpha(0,\, t)\rangle\end{aligned} \tag{40.46}$$

Here, $j_\beta(\mathbf{r},\, t')$ is scalarly multiplied by $E_\beta(t')$.

If E_β depends upon time according to the harmonic law $E_\beta(t) = E_{0\beta}e^{-i\omega t}$, then

$$\langle \hat{j}_\alpha(t)\rangle = E_{0\beta}e^{-i\omega t}\int_{-\infty}^{t} dt'\, e^{-i\omega(t'-t)} \times \int dV \langle j_\beta(\mathbf{r},\, t'-t+t)\, j_\alpha(0,\, t)\rangle \qquad (40.47)$$

Now, substituting $t - t' \equiv \tau$, we obtain

$$\langle j_\alpha(t)\rangle = \frac{E_{0\beta}}{\theta}\, e^{-i\omega t}\int_0^{\infty} d\tau e^{i\omega\tau}\int \langle j_\beta(\mathbf{r},\, t+\tau)\, j_\alpha(0,\, t)\rangle\, dV \qquad (40.48)$$

The mean value over the equilibrium state does not depend upon time, hence we can put $t = 0$ in j_β, j_α without restricting the generality. Comparing the obtained expression (40.48) with (40.33) and (40.34), we see that the Kubo formula has been proved.

Similar expressions can be obtained for other kinetic coefficients. Thus, the relationships between linear transport coefficients are found with the help of equilibrium distribution functions. The possibility of such a description is apparent from the fact that both equilibrium systems and systems weakly deflected from equilibrium are described by the same Hamiltonian. The perturbation introduced by an external action is expressed in terms of the Hamiltonian and equilibrium distribution function linearly with respect to the perturbation energy $\hat{V}$ (40.41).

EXERCISES

1. Show that the static electric conductivity tensor is symmetric.

Solution. Rewrite formula (40.13) for a homogeneous anisotropic medium:

$$\frac{dS}{dt} = \int \frac{E_\alpha j_\alpha}{\theta}\, dV$$

where α is the tensor index. By analogy with (40.15), we conclude from this that

$$X_\alpha = -E_\alpha/\theta$$

Since $j_\alpha = \sigma_{\alpha\beta}E_\beta \equiv -\theta\sigma_{\alpha\beta}X_\beta$, from the Onsager reciprocity theorem we obtain the required symmetry relationship $\sigma_{\alpha\beta} = \sigma_{\beta\alpha}$.

2. Show that the electric conductivity tensor in a rapidly variable field is symmetric.

Solution. From formulas (40.33) and (40.34) we obtain

$$\sigma_{\alpha\beta}(\omega)=\frac{1}{\theta}\int\limits_0^\infty \sigma_{\alpha\beta}(\tau)\, e^{i\omega\tau}\, d\tau$$

$$=\frac{1}{\theta}\int\limits_0^\infty e^{i\omega\tau}\, d\tau \int dV\, \langle j_\alpha(\mathbf{r},\ \tau)\, j_\alpha(0,\ 0)\rangle$$

From the time symmetry of fluctuations we conclude that

$$\langle j_\beta(\mathbf{r},\ \tau)\, j_\alpha(0,\ 0)\rangle = \langle j_\beta(\mathbf{r},\ 0)\, j_\alpha(0,\ \tau)\rangle$$

Now moving the origin of the coordinate system to point $\mathbf{r}$, and taking into account that substitution of $-\mathbf{r}$ for $\mathbf{r}$ changes nothing in integration over the volume, we arrive at the equation

$$\int dV\, \langle j_\beta(\mathbf{r},\ \tau)\, j_\alpha(0,\ 0)\rangle=\int dV\, \langle j_\alpha(-\mathbf{r},\ \tau)\, j_\beta(0,\ 0)\rangle$$

which proves the symmetry of the conductivity tensor.

3. Using the results of Exercise 4, Section 17, express the diffusion coefficient in terms of the correlation between the velocities of a particle at different instants.

Solution. We proceed from the identity

$$\mathbf{r}=\int\limits_0^t \mathbf{v}(t')\, dt'$$

and find the square of $\mathbf{r}$:

$$r^2=\int\limits_0^t \mathbf{v}(t')\, dt' \int\limits_0^t \mathbf{v}(t'')\, dt''=\int\limits_0^t dt' \int\limits_0^t dt''\mathbf{v}(t')\, \mathbf{v}(t'')$$

Now assume that at some initial time N particles are emitted from an arbitrary point (N being a large number). Each particle experiences random collisions with molecules of the medium independently of the other emitted particles. At time t a particle i will be at a distance r_i from the initial point. Averaging r^2 over the diffusing particles, we get

$$\overline{r^2}=\frac{1}{N}\sum_{i=1}^{N} r_i^2=\frac{1}{N}\int\limits_0^t dt' \int\limits_0^t dt'' \sum_{i=1}^{N} \mathbf{v}_i(t')\, \mathbf{v}_i(t'')$$

$$=\int\limits_0^t dt' \int\limits_0^t dt''\, \overline{\mathbf{v}(t')\, \mathbf{v}(t'')}$$

and substitute $t'+\tau$ for t''. At large time intervals τ the velocity correlation is lost due to the chaotic nature of the collisions: a particle "forgets" its velocity at the time t'. Therefore $\lim \overline{\mathbf{v}(t')\, \mathbf{v}(t'+\tau)} = 0$ as $\tau \to \pm\infty$.

Taking this into account, we can rewrite the double integral as follows:

$$\overline{r^2} = \int_0^t dt' \int_{-\infty}^{\infty} \overline{\mathbf{v}(t')\,\mathbf{v}(t'+\tau)}\,d\tau$$

But by virtue of time homogeneity the internal integral over τ cannot depend on t'. Therefore

$$\overline{r^2(t)} = t \int_{-\infty}^{\infty} \overline{\mathbf{v}(0)\,\mathbf{v}(\tau)}\,d\tau = 6Dt$$

whence

$$D = \frac{1}{6} \int_{-\infty}^{\infty} \overline{\mathbf{v}(0)\,\mathbf{v}(\tau)}\,d\tau = \frac{1}{3} \int_{0}^{\infty} \overline{\mathbf{v}(0)\,\mathbf{v}(\tau)}\,d\tau$$

This relationship is in effect a special case of the Kubo formula with the conventional mean substituted for the quantum mechanical. The quantity D is also a transport coefficient connected with the mobility of the particles by the Einstein relationship (17.26). Conductivity, in turn, is expressed in terms of mobility as follows: $\sigma = ne\tilde{\omega}$, where n is the number of charge carriers per unit volume.

41

THE TRANSPORT EQUATION

The relationships obtained in the preceding section are analogous to thermodynamic relationships: they do not depend on the properties of the specific medium and merely interrelate certain mean quantities. It does not follow from the formulas how transport coefficients can be expressed in terms of atomic constants.

In the most general case, for specific computations one must know the density matrix of the nonequilibrium state (40.35). It provides the fullest possible description of a system subject to external action and in contact with the surroundings. At small deviations from equilibrium the density matrix is given by formula (40.41), which, however, is too general to be used for obtaining results referring to specific systems.

The diagonal elements of the density matrix taken at $x = x_1'$ give the distribution function of the system we employed in the statistics of equilibrium systems. But the distribution-function

concept can also be used with respect to nonequilibrium systems, if the off-diagonal elements of $\rho(x', x)$ are of no great consequence in the problem concerned. In this section we shall show how to find the distribution function directly, bypassing the density matrix. The distribution function, computed to a greater or lesser degree of accuracy, can then be used to determine such transport coefficients as viscosity, heat conductivity, etc.

For this, one must first construct equations expressing the balance of the number of particles passing from one state of the system to another. The accuracy of such equations depends on the exactness to which the state of each particle is defined and on how well the probability of transition between different states is represented. The equations for nonequilibrium distribution functions are called *transport equations.*

The Diffusion Coefficient. Let us consider the example of a transport equation in which the simplifications introduced into the problem in advance are clearly seen. Let us show how to determine the distribution function describing particles diffusing in a certain medium. They fill the volume of the medium nonhomogeneously, their concentration n being dependent on the coordinates. In the absence of external forces sustaining nonhomogeneity, such as gravity, the particles are not in statistical equilibrium. A velocity distribution establishes between them that results in their equilibrium, that is, in a uniform velocity distribution in space. Let us show how to find the velocity distribution function for this case.

If we direct the x axis along the concentration gradient, the distribution function will be dependent on x and on the particle's velocity $\mathbf{v}$, that is, $n = n(x, \mathbf{v})$. But we introduce a simplification into the problem and concern ourselves only with the direction of the velocity and not its absolute value; thus, we shall seek $n(x, \mathbf{v}_0)$, where $\mathbf{v}_0$ is a unit vector in the direction of $\mathbf{v}$. We put the absolute values of $\mathbf{v}$ the same for all particles. The basic aspects of the phenomenon of diffusion are detectable in such a rough description.

We shall call the probability that in unit time a certain particle will change the direction of the velocity from $\mathbf{v}_0$ to $\mathbf{v}_0'$ as follows:

$$dW_0 = w(|\mathbf{v}_0' - \mathbf{v}_0|)\, d\Omega' \tag{41.1}$$

In Section 17 we considered the diffusion of small but macroscopic particles. If the particles—atoms or molecules—diffuse in a gas, their interactions with the medium are of the nature of individual collisions. Then dW_0 is expressed in terms of the effective cross section.

We assume the medium to be isotropic, so that probability depends only on the angle between $\mathbf{v}_0'$ and $\mathbf{v}_0$, and not on their direc-

tion in space. The solid angle element $d\Omega'$ is taken in a coordinate system in which the components of $\mathbf{v}'$ are laid off along the axes.

The total change in the number of particles with the given vector $\mathbf{v}_0$ in unit time and unit volume is zero (taking into account all particles with vector $\mathbf{v}_0$ entering and leaving the volume):

$$\frac{d}{dt} n(t, x, \mathbf{v}_0) = 0$$

We present this equation in explicit form. The partial derivatives with respect to the first two variables are computed in the usual way. The direction of the velocity $\mathbf{v}_0$ can change precipitously, for instance in the collision of a given particle with a particle of the medium (mutual collisions of the diffusing particles are not considered here). The change in the number of particles with the given direction of the velocity $\mathbf{v}_0$ is compounded of two parts: the first is taken with the minus sign and expresses the decrease in the number of such particles due to collisions

$$-\int n(\mathbf{v}_0)\, dW_0 = -\int n(\mathbf{v}_0)\, w(|\mathbf{v}_0' - \mathbf{v}_0|)\, d\Omega' \tag{41.2}$$

the second (with the plus sign) expresses the passage of particles into the state with the given $\mathbf{v}_0$, from states with all different directions of velocity $\mathbf{v}_0$:

$$\int n(\mathbf{v}_0')\, dW_0 = \int n(\mathbf{v}_0')\, w(|\mathbf{v}_0' - \mathbf{v}_0|)\, d\Omega' \tag{41.3}$$

Thus we arrive at the following balance equation:

$$\frac{dn}{dt} = \frac{\partial n}{\partial t} + \frac{\partial n}{\partial x}\frac{dx}{dt} + \int [n(\mathbf{v}_0) - n(\mathbf{v}_0')]\, w(|\mathbf{v}_0' - \mathbf{v}_0|)\, d\Omega' = 0 \tag{41.4}$$

In steady-state conditions $\partial n/\partial t = 0$. Denote by ϑ the angle between the velocity and the x axis. Then $dx/dt = v\cos\vartheta$, and (41.4) acquires the form

$$v\cos\vartheta\,\frac{\partial n}{\partial x} = -\int [n(\mathbf{v}_0) - n(\mathbf{v}_0')]\, w(|\mathbf{v}_0' - \mathbf{v}_0|)\, d\Omega' \tag{41.5}$$

According to our simplification, here v = constant. We can choose it equal to, say, the mean thermal velocity.

We represent the distribution function in the following form:

$$n(x, \mathbf{v}_0) = n_0(x) + n_1(x, \cos\vartheta) \tag{41.6}$$

Assuming the deviation from equilibrium to be small, we must require a strong inequality $n_0 \gg n_1$. Then only n_0 remains in the left-hand side of Eq. (41.5), and only n_1 in the right-hand side. This is

readily understood if we take into account that $n_0(x)$ does not depend on $\mathbf{v}_0$, and that if $n_0(x)$ is substituted into the integrand, the latter becomes zero. Thereby the problem reduces to a nonhomogeneous linear integral equation

$$v \cos\vartheta \frac{\partial n_0}{\partial x} = -\int w(|\mathbf{v}_0' - \mathbf{v}_0|)\,[n_1(x, \cos\vartheta) - n_1(x, \cos\vartheta')]\,d\Omega' \tag{41.7}$$

We separate the angular dependence in the following way:

$$n_1(x, \cos\vartheta) = n_1(x)\cos\vartheta \tag{41.8}$$

Accordingly, $n_1(x, \cos\vartheta) = n_1(x)\cos\vartheta$. In Figure 53 are shown vectors $\mathbf{v}_0$, $\mathbf{v}_0'$ and the angles between them. From the definition of

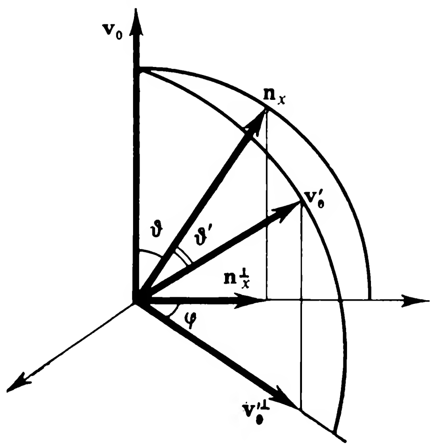


Figure 53

ϑ we can write $\cos\vartheta' = (\mathbf{v}_0' \cdot \mathbf{n}_x)$, where $\mathbf{n}_x$ is a unit vector along the x axis). We resolve this dot product into two terms formed by the component vectors perpendicular and parallel to $\mathbf{v}_0$:

$$\cos\vartheta' \equiv (\mathbf{v}_0' \cdot \mathbf{n}_x) = v_0'^{\|} n_x^{\|} + (\mathbf{v}_0'^{\perp} \cdot \mathbf{n}_x^{\perp})$$

If θ is the angle between $\mathbf{v}_0$ and $\mathbf{v}_0'$, then $v_0'^{\|} = \cos\theta$, $n_x^{\|} = \cos\vartheta$, $v_0'^{\perp} = \sin\theta$, and $n_x^{\perp} = \sin\vartheta$. Introducing the angle between vectors $\mathbf{v}_0'^{\perp}$ and $\mathbf{n}_x^{\perp}$ in the plane, φ, we obtain

$$\cos\vartheta' = \cos\vartheta\cos\theta + \sin\vartheta\sin\theta\cos\varphi \tag{41.9}$$

(This is a fundamental formula of spherical trigonometry, in which triangles are considered not in a plane but on a sphere of unit ra-

dius.) The integration in (41.7) is over the whole solid angle. Therefore we can take $\mathbf{v}_0$ for the polar axis, substituting $\sin\theta\, d\theta\, d\varphi$ for $d\Omega'$. Substituting (41.9) into (41.7), we get

$$v\cos\vartheta\frac{\partial n_0}{\partial x} = -2\pi n_1(x)\cos\vartheta\int_0^\pi w_0(\cos\theta)(1-\cos\theta)\sin\theta\, d\theta$$

(integration over φ reduces the term involving $\cos\varphi$ to zero[2]). Now we determine the following quantity:

$$l_t \equiv \frac{v}{2\pi}\left[\int_0^\pi w_0(\cos\theta)(1-\cos\theta)\sin\theta\, d\theta\right]^{-1}$$

$$= \frac{v}{\int w_0(1-\cos\theta)\, d\Omega} \tag{41.10}$$

(the meaning of the symbol l_t will be explained later on). Then the required correction to the distribution function is

$$n_1(x) = -l_t\frac{\partial n_0}{\partial x} \tag{41.11}$$

Since the term $n_0(x)$ makes no contribution to the flux of particles, that is, $\int \cos\vartheta\, n_0\, d\Omega = 0$, the required diffusion flux is found as follows:

$$j = \int v\cos\vartheta n\, d\Omega = -\frac{\partial n_0}{\partial x}vl_t\int \cos^2\vartheta\, d\Omega$$

$$= -\frac{4\pi}{3}vl_t\frac{\partial n_0}{\partial x} \tag{41.12}$$

Comparing (41.12) with the general expression (17.21), which links the diffusion flux with the concentration gradient, we find the diffusion coefficient ($4\pi n_0 = \int n\, d\Omega$ is the volume concentration)

$$D = \frac{1}{3}vl_t \tag{41.13}$$

Length of Free Path. If individual atoms or molecules are diffusing the obtained expression can be interpreted in the following way. In scattering, the probability of a particle deflecting by an angle θ in unit time is expressed in terms of the effective cross section $d\sigma$

---

[2] In the theory of spherical functions it is proved that

$$\int_0^{2\pi} d\varphi\, P_n(\cos\vartheta') = 2\pi P_n(\cos\vartheta)\cdot P_n(\cos\theta)$$

where P_n is the nth Legendre polynomial. Here the equation is proved for $n = 1$.

[Secs. 5 and 35]:

$$dW_0 = vN\,d\sigma \tag{41.14}$$

Here, when we speak of collisions of atoms, $d\sigma$ has a quite definite meaning.[3]

The transport equation (41.5) can be written with the help of the probability determined by formula (41.14). Then the quantity l_t, formerly expressed by definition (41.10), is written as

$$l_t = \left[\int N\,d\sigma\,(1-\cos\theta) \right]^{-1} \tag{41.15}$$

From this follows the correlation between two ways of stating scattering probability:

$$vN\,d\sigma = w\,d\Omega \tag{41.16}$$

We shall now show how the obtained results correspond to the elementary notions of transport processes.

As was pointed out in Section 35, the total effective scattering cross section may converge if the forces between two particles falls off fast enough with distance. In the nonquantum theory of scattering this requires the forces to become identically zero at some distance from the scattering centre. Suppose the effective scattering cross section is finite and equal to σ. If the impinging particles form a parallel beam of flux density I, its decrease per unit path is

$$\frac{dI}{dx} = -N\sigma I$$

or

$$I = I_0 e^{-N\sigma x} \tag{41.17}$$

On the other hand, in the kinetic theory of gases the concept of free path l of a particle between collisions is introduced. The weakening of a parallel beam expressed in terms of l has the form of a damping exponential function $I = I_0 e^{-x/l}$. Comparing this with (41.17), we find the connection between σ and l:

$$l = (N\sigma)^{-1} = \frac{1}{\int N\,d\sigma} \tag{41.18}$$

The length l_t involved in the diffusion coefficient is expressed somewhat differently, as can be seen from (41.15). It is called the *transport path*, because the transport of particles is defined in terms of it. The length l_t coincides with l only in isotropic scattering, when $\int \cos\theta\,d\sigma = 0$.

[3] If in the collisions of molecules we ignore their spatial orientations, formula (41.14) is sufficient in the adopted approximation.

If scattering occurs with greater probability at $\theta \ll \pi/2$, that is, forward, the main contribution to the integral defining l_t^* is at $\cos\theta \approx 1$. The integral in the denominator of (41.15) can become much smaller than $N\sigma$; then $l_t \gg l$. In these conditions, from (41.13), the diffusion coefficient is substantially greater than yielded by the elementary evaluation in which l is substituted for l_t.

The inequality $n_1 \ll n_0$ was taken as the basis of the computations. Its meaning is apparent from formula (41.11): the inequality is satisfied if the particle density varies but slightly along the transport path. A similar condition applies not only to diffusion theory but to other transport phenomena as well: heat conductivity (energy transfer), viscosity (momentum transport). In most nonequilibrium processes it is satisfied.

But in shock compression, for example, if the density changes greatly, the whole irreversible process takes place along a single path length. In this case the very terms "heat conductivity" and "viscosity" are inapplicable.

In developing the diffusion coefficient we did not take account of the velocity distribution of the diffusing particles. Owing to this it was meaningless to introduce formulas showing the dependence of the effective cross section on the magnitude of the relative velocity.

The factor 1/3 in the expression for D is written only because it was obtained in computations. The assumption that the velocities of all diffusing particles are the same is in any case of a qualitative character.

The Dependence $\sigma(v)$. For subsequent applications we shall find the dependence of the classical effective diffusion cross section upon the velocity. Suppose the forces decrease with distance according to the power law $F = ar^{-n-1}$. The dimension of the factor a is ml^{n+2}/t^2. The effective scattering cross section can depend only on a, mass, and velocity. But its expression is determined from the dimensions in only one way:

$$\sigma \sim \left(\frac{a}{mv^2}\right)^{2/n} \sim \frac{1}{v^{4/n}}, \qquad [\sigma] = l^2$$

Consequently, the probability of collision in unit time is proportional to the velocity in the $(1 - 4/n)$th power. Obviously, if $n = 4$, the scattering cross section σ does not depend on the velocity. Such a power for repulsion forces was introduced by Maxwell mainly for the convenience of calculations.

Nevertheless, an ionized atom interacts with a neutral one precisely according to Maxwell's law, but with the sign of the force reversed: they attract, not repulse, as Maxwell had assumed in his model of interaction of neutral atoms. The energy of a neutral atom in an

ion's field is **Ed**, where **d** is the induced dipole moment. If the atom's polarizability is α, then $\mathbf{d} = \alpha\mathbf{E}$, so that the interaction energy is proportional to the square of the field. The field of an ion is, in turn, inversely proportional to the square of the distance, so that the potential energy of interaction is proportional to r^{-4}, and the force varies in proportion to r^{-5}. This result will be utilized in Exercise 3.

The Boltzmann Transport Equation. We shall now consider an equation that makes it possible to determine the nonequilibrium distribution function in finer detail: not only according to the directions but according to the magnitude of the velocity as well. In the strict sense the equation will refer to a monatomic gas, in other words it will take into account only the transport degrees of freedom of colliding particles. If the gas is not very dense, it is sufficient to consider only collisions of pairs of atoms, as was done in Section 11.

Let us obtain the expression involved in the balance of collisions analogous to (41.2) and (41.3). Let the required distribution function of the atoms with respect to the coordinates and velocities be $f(t, \mathbf{r}, \mathbf{v})$. We shall determine its variation at a given point in space due to collisions with atoms having velocity $\mathbf{v}'$. If we denote the number of atoms in unit volume by n (in the preceding item it was more convenient to take the density $4\pi n$), the number of atoms having the velocity $\mathbf{v}$ in unit volume will be $nf(t, \mathbf{r}, \mathbf{v})$. In other words, the function $f(t, \mathbf{r}, \mathbf{v})$ is normalized to unity. The number of collisions of an atom having velocity $\mathbf{v}$ with all atoms having velocity $\mathbf{v}'$ in unit time is equal to $nf(t, \mathbf{r}, \mathbf{v}')(\mathbf{v} - \mathbf{v}')\,d\sigma$. Note the following relationships: the relative velocity of the pair, $|\mathbf{v} - \mathbf{v}'|$, corresponds to v in (41.14); the number of atoms in unit volume colliding with the given atom, $nf(t, \mathbf{r}, \mathbf{v}')$, is analogous to N. Then the change in the distribution in unit time due to such collisions is, as in (41.2), equal to

$$-f(t, \mathbf{r}, \mathbf{v}) \int\int nf(t, \mathbf{r}, \mathbf{v}')\,|\mathbf{v} - \mathbf{v}'|\,d\sigma\,d\tau_{\mathbf{v}'} \qquad (41.19a)$$

Let the velocities of the atoms as a result of the collision become $\mathbf{v}_1$ and $\mathbf{v}_1'$. Assuming all the collisions to be elastic, that is, not accompanied by electron excitation, we find that the spatial direction of the relative velocity of the atoms is affected, but not the absolute value [Sec. 6]:

$$|\mathbf{v} - \mathbf{v}'| = |\mathbf{v}_1 - \mathbf{v}_1'| \qquad (41.20)$$

From the principle of detailed balance, the probabilities of a direct and reverse collision in unit time are the same (Sec. 1). They derive one from the other by changing the sign of time in the equations

of mechanics invariant with respect to the operation $t \to -t$. Therefore, the change in the distribution function $f(t, \mathbf{r}, \mathbf{v})$ as a result of the collisions (where $\mathbf{v}$ is the final velocity in unit time) is, similarly to (4.19), equal to

$$\int\int nf(t, \mathbf{r}, \mathbf{v}_1) f(t, \mathbf{r}, \mathbf{v}_1') \, |\mathbf{v}_1 - \mathbf{v}_1'| \, d\sigma \, d\tau_{\mathbf{v}'} \qquad (41.19b)$$

The difference between the integrals (41.19*a*) and (41.19*b*) is the required balance of collisions. The angle of turn of the unit vector of relative velocity in a reference system connected with the centre of mass is equal to $(\mathbf{v}-\mathbf{v}')/(|\mathbf{v}-\mathbf{v}'|)$. This angle, denoted χ, on which $d\sigma$ depends, uniquely connects $\mathbf{v}$ and $\mathbf{v}'$ with $\mathbf{v}_1$ and $\mathbf{v}_1'$. Going over from $d\sigma$ to χ, we obtain the transition formulas from $\mathbf{v}$, $\mathbf{v}'$ to $\mathbf{v}_1$, $\mathbf{v}_1'$ (see Exercise 3, also [6.8] and [6.9]). The transport equation for $f(t_1, \mathbf{r}_1, \mathbf{v})$ has the following form:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \left(\frac{d\mathbf{r}}{dt} \cdot \operatorname{grad} f\right) + \frac{d\mathbf{v}}{dt}\frac{\partial f}{\partial \mathbf{v}}$$
$$+ \int\int n\,[f(t, \mathbf{r}, \mathbf{v}) f(t, \mathbf{r}, \mathbf{v}') - f(t, \mathbf{r}, \mathbf{v}_1) f(t, \mathbf{r}, \mathbf{v}_1')]$$
$$\times |\mathbf{v}-\mathbf{v}'| \frac{d\sigma}{d\Omega_\chi} \, d\Omega_\chi \, d\tau_{\mathbf{v}'} \qquad (41.21)$$

Here in place of $d\mathbf{v}/dt$ we must substitute $\mathbf{F}/m$ (where $\mathbf{F}$ is the force of the external field acting on the atom), and $\mathbf{v}$ must be substituted for $d\mathbf{r}/dt$.

Unlike (41.4), Eq. (41.21) is nonlinear, because the state is established as a result of collisions between atoms of the same gas. In developing (41.4) it was assumed that the state of the medium in which diffusion takes place is not changed by the action of the diffusing particles.

If the gas is homogeneous and not situated in an external field, that is, $\operatorname{grad} f = 0$ and $\mathbf{F} = 0$, the steady state corresponding to the condition $\partial f/\partial t = 0$ is given by the Maxwell distribution $f(v) = \exp[-mv^2/(2\theta)]$. In any case this distribution satisfies (41.21), since by virtue of the energy conservation law

$$f(v) f(v') - f(v_1) f(v_1') = 0$$

In a uniform external field we have the Boltzmann distribution

$$f = \exp\left[-\frac{1}{\theta}\left(\frac{mv^2}{2} + U\right)\right]$$

because

$$\operatorname{grad} f = -\frac{\operatorname{grad} U}{\theta} f = \frac{\mathbf{F}f}{\theta}, \qquad \frac{\partial f}{\partial \mathbf{v}} = -\frac{\mathbf{v}}{\theta} f$$

$$(\mathbf{v} \cdot \operatorname{grad} f) + \frac{d\mathbf{v}}{dt}\frac{\partial f}{\partial \mathbf{v}} = 0$$

Consequently, both these distribution functions assure the equilibrium state of a system.

The H Theorem. Boltzmann showed with the help of the transport equation that, as a result of collisions among the atoms, the initial nonequilibrium state of a gas moves to equilibrium, or at least to a steady state at which the balance of the number of collisions becomes zero.

For this Boltzmann introduced the *H function* (not to be confused with enthalpy) analogous to entropy:

$$H \equiv -\int d\tau_{\mathbf{v}} f(v, t) \ln f(v, t) \tag{41.22}$$

Let us calculate the derivative of H with respect to time:

$$\frac{\partial H}{\partial t} = \int d\tau_{\mathbf{v}} \frac{\partial f}{\partial t} \ln f - \int d\tau_{\mathbf{v}} \frac{\partial f}{\partial t}$$

$$= -\int d\tau_{\mathbf{v}} \frac{\partial f}{\partial t} \ln f - \frac{\partial}{\partial t} \int f \, d\tau_{\mathbf{v}}$$

From the normalization condition the time derivative of the last integral is zero. We replace the partial derivative under the integration sign according to Eq. (41.21):

$$\frac{\partial H}{\partial t} = n \int d\tau_{\mathbf{v}} \int d\tau_{\mathbf{v}'} \int d\sigma \, |\mathbf{v} - \mathbf{v}'| \ln f(\mathbf{v}) \times [f(\mathbf{v}) f(\mathbf{v}') - f(\mathbf{v}_1) f(\mathbf{v}_1')] \tag{41.23}$$

This expression is symmetrical with respect to the substitution of $\mathbf{v}'$ for $\mathbf{v}$ since both velocities are integration variables. Besides, $\mathbf{v}_1$ and $\mathbf{v}_1'$ can be substituted for $\mathbf{v}$ and $\mathbf{v}'$, provided the sign of the integral is changed. Finally, the integral is not affected by a permutation between $\mathbf{v}_1$ and $\mathbf{v}_1'$. We now symmetrize (41.23) with respect to all four velocities and multiply the result by 1/4 to get

$$\frac{\partial H}{\partial t} = \frac{n}{4} \int d\tau_{\mathbf{v}} \int d\tau_{\mathbf{v}'} \int d\sigma \, |\mathbf{v} - \mathbf{v}'| \ln \frac{f(\mathbf{v}) f(\mathbf{v}')}{f(\mathbf{v}_1) f(\mathbf{v}_1')} \times [f(\mathbf{v}) f(\mathbf{v}') - f(\mathbf{v}_1) f(\mathbf{v}_1')] \tag{41.24}$$

A logarithm is a monotonic function of its argument. Therefore, an expression of the form $(x - y) \ln x/y$ is positive both at $x \geqslant y$ and $x < y$. It follows from this that H can only increase, that is, $\partial H/\partial t \geqslant 0$. The quantity H attains its maximum when the expressions in brackets become zero. But then, as we have seen, $f(\mathbf{v})$ coincides with the Maxwell distribution.

If we assume that the integral involved in (41.24) does not become zero in the substitution of other distribution functions besides the Maxwellian, it follows from this equation that the nonequilibrium velocity distribution of the collisions among atoms will turn the

nonequilibrium velocity distribution of the atoms into an equilibrium distribution. In this way Boltzmann substantiated the principle of increasing entropy and the related establishment of equilibrium. Unfortunately, this did not convince many of his contemporaries, who regarded the atom as more of a speculative concept than a physical reality: at the time there was no direct experimental proof of the existence of atoms.

The Boltzmann transport equations are based on the laws of mechanics, which are symmetrical with respect to time inversion. The equation itself is of the first order with respect to time and asymmetrical. This is seen especially well from the *H theorem*, which leads only to an increase in entropy. This is due to the nature of the very posing of the problem in kinetics: the nonequilibrium state is treated as the initial state.

Spontaneous deviations from equilibrium, or fluctuations, are not covered by Eq. (41.21). For instance, in developing it the mean number of collisions undergone by the atoms in unit time, $n(\mathbf{v} - \mathbf{v}') f(\mathbf{v}') \, d\sigma$, is substituted. This approach does not take account of fluctuations, as can be seen from (41.24). But as was shown in Section 10, in statistics it is precisely fluctuations that restore the symmetry with respect to the direction of time.

Relaxation Time. Solution of the Boltzmann nonlinear integroe differential equation for the most general case presents formidabl- difficulties. Such a solution may be required to study the state of a gas in a strong shock front. Note that this problem is also of practical importance. At supersonic flight of missiles or satellites in the upper, sufficiently rarefied, layers of the atmosphere the free path of a molecule is comparable with the dimensions of the moving body. But the width of the shock front in which the pressure doubles or more is a quantity of the same order as the free path. Consequently, the air flowing around the body has a nonequilibrium distribution function throughout the flow domain. Here a hydrodynamic description is impossible. One must therefore know the nonequilibrium distribution function in some approximation.

When the deviations from a statistical distribution are small, the problem nevertheless permits a general investigation, since it can be linearized. Suppose that the factors causing the deviation from equilibrium are given by quantities of the first order. These may be the velocity or temperature gradients or external forces due to the absence of statistical equilibrium (for example, electric fields in conductors). All such factors are described by the second and third terms in the transport equation (41.21). We represent the required distribution function as follows:

$$f_0 [1 + g(\mathbf{v}, t)] \tag{41.25}$$

where f_0 is the Maxwell distribution function, and g is a function of the first order with respect to the disturbances. That is why in the transport equation we should neglect any product of quantities of the same order, and any products of functions of different arguments.

As was shown, the function f_0 identically satisfies the transport equation. Therefore terms linear with respect to g should be left under the integral sign.

The derivative $\partial f/\partial t$ can be substituted by $f_0 (\partial g/\partial t)$. The second and third terms are proportional to the magnitudes of the disturbing factors. They make the equation nonhomogeneous with respect to g. As always, it is useful to first consider a homogeneous equation. If under the integral sign we substitute $f_0(\mathbf{v}) f_0(\mathbf{v}')$ for $f_0(\mathbf{v}_1) f_0(\mathbf{v}_1')$, after cancelling out $f_0(\mathbf{v})$ we arrive at a linear homogeneous integro-differential equation with respect to the correction function:

$$\frac{\partial g}{\partial t} = -n \int\int |\mathbf{v}-\mathbf{v}'|\, d\tau_{\mathbf{v}'}\, d\sigma\, f_0(\mathbf{v}') \times [g(\mathbf{v}) + g(\mathbf{v}') - g(\mathbf{v}_1) - g(\mathbf{v}_1')] \quad (41.26)$$

It can be seen from this equation that the integral operator applied to $g(\mathbf{v})$ has the dimension $1/\tau$ (where τ is a certain characteristic time). We seek the solution of (41.26) in the form

$$g = g_0(\mathbf{v})\, e^{-t/\tau}$$

that is, with separated variables. Then the function $g(t)$ satisfies the equation

$$\frac{g_0}{\tau} = \hat{L} g_0 \quad (41.27)$$

where $\hat{L}$ is the same linear integral operator as in (41.26). From the notation of Eq. (41.27) we see that $1/\tau$ is the eigenvalue of this operator. It can be proved that the operator $\hat{L}$ is Hermitian, which implies that it has real eigenvalues [Sec. 25] and, that, in addition, all real eigenvalues of $\hat{L}$ are positive. Simple integral transformations are used for this.

Further, in Eq. (41.27) the angular dependence $g_0(\mathbf{v})$ is separated. We seek $g_0(\mathbf{v})$ in the form

$$g_0(\mathbf{v}) = G_l(v)\, P_l^m(\cos\vartheta)\, e^{im\varphi} = G_l(v)\, Y_l^m \quad (41.28)$$

where P_l^m is an associated Legendre polynomial, and Y_l^m is a spherical function.

Since Eq. (41.27) does not change its form in any spatial rotation of the coordinate system (there is no preferred direction), substitution of $g(\mathbf{v}')$, $g(\mathbf{v})$, $g(\mathbf{v}_1)$, $g(\mathbf{v}_1')$ with (41.28) in mind yields, after integration over the angles, an expression proportional to the same spherical function that was substituted. For example, substitution

of the expression $g(\mathbf{v}')$, which involves the spherical function of the angles in $\mathbf{v}'$-space, yields the same function in $\mathbf{v}$-space; the result is similar in the case of $\mathbf{v}_1$ and $\mathbf{v}_1'$. An example of such substitution is presented by integration of the function in (41.8). The scalar operator $\hat{L}$ applied to a spherical function can yield only that same function. Symbolically we write this down in the following form:

$$\hat{L}g = \hat{L}G_l(v') Y_l^m(\vartheta', \varphi') = (\hat{\Lambda}_l G_l(v') Y_l^m(\vartheta, \varphi)) \qquad (41.29a)$$

Here, $\hat{\Lambda}_l$ is a scalar operator depending only upon the absolute magnitude of the velocity. It is determined by the concrete differential effective cross section $d\sigma/d\Omega_\chi$ and the order of the spherical function of l but not by the number m, which depends only upon an arbitrary choice of the polar axis in space.

As pointed out before, after separation of the angular dependence in $g(\mathbf{v})$ there still remains the dependence upon the absolute magnitude of the velocity, just as in the Schrödinger equation after separating the variables there remains in the central field an equation for the radial function $\varphi(r)$. The energy eigenvalue is determined by two numbers, the azimuthal l and the radial n_r, that is, the number of nodal points in $\varphi(r)$. The number l is involved in the equation of the present problem as well, but instead of n_r there is a certain number s representing the eigenfunction of the equation involving only the absolute magnitude of the velocity. Denoting this function $G_{ls}(v)$, we find that its eigenvalue equation has the form

$$\hat{\Lambda}_l G_{ls} = \frac{1}{\tau_{ls}} G_{ls} \qquad (41.29b)$$

The quantity τ_{ls} has the dimensions of time and is called the *relaxation time* of the system with a nonequilibrium distribution function $g_{ls}(\mathbf{v}) = G_{ls}Y_l^m$. It gives the time in which the correction to the equilibrium function falls off by a factor of e. The relaxation time thus obtained is not an estimate but a quite precise quantity if the distribution functions are taken as eigenfunctions with respect to the collision operator $\hat{L}$.

There is a wide range of relaxation times τ_{ls}. Different l and s correspond to different processes.

Viscosity of a Monatomic Gas. In elementary courses of physics, transport coefficients (viscosity or heat conductivity of a gas) are defined in terms of the free path of a molecule or atom. We assume these evaluations to be known to the reader. Here we shall show how to compute the viscosity coefficient of a monatomic gas from the transport equation. Comparison with experimental data in principle allows for a reconstruction of the elementary law of strong interactions between atoms according to the temperature dependence of the viscosity coefficient.

Consider a gas with a given uniform mean-velocity field. Let the mean velocity component along the x axis be linearly dependent upon y: $v_x = \alpha y$. In other words, at a point at a distance y from the plane $y = 0$ the velocity of the centre of mass of an elementary volume of the gas is proportional to y. We assume that at $y = 0$ there is a solid fixed wall. Then v_x is the xth component of the hydrodynamic velocity of the gas relative the wall.

The Maxwell distribution in an elemental volume is described by the following exponential function:

$$f_0 = \left(\frac{m}{2\pi\theta}\right)^{3/2} e^{-m/(2\theta)\,[(v_x - \alpha y)^2 + v_y^2 + v_z^2]} \tag{41.30}$$

But such a distribution does not satisfy the transport equation (41.21): the integral operator becomes zero while the term $\partial f_0/\partial y$ is not zero. Consequently the distribution function for such conditions should be sought in the form

$$f = f_0\,[1 + g\,(v_x - \alpha y,\ v_y,\ v_z)] \tag{41.31}$$

The correction function g need not be taken into account in the term $\partial f_0/\partial y$ since f_0 is already dependent on y (and the dependence is assumed weak). Therefore the product $(\partial f_0/\partial y)g$ should be neglected. The equation for g assumes the form

$$v_y \frac{\partial f_0}{\partial y} + f_0 \hat{L} g = 0 \tag{41.32}$$

Here the operator $\hat{L}$ is given by Eq. (41.27), from which it follows that $\hat{L}g = g/\tau$. The exponential factor $e^{-t/\tau}$ changes nothing in (41.27)

From (41.25), $f_0 g = f - f_0$; therefore the transport equation reduces to the form

$$v_y \frac{\partial f_0}{\partial y} + \frac{f - f_0}{\tau} = 0 \tag{41.33}$$

This is the *relaxation equation*, and the approximation is called the *relaxation-time approximation.* To give it strict meaning the appropriate value of τ must be substituted into it. Let us show the condition from which τ can be found. The derivative $\partial f_0/\partial y$ is expressed as follows:

$$\frac{\partial f_0}{\partial y} = \frac{m\,(v_x - \alpha y)\,f_0}{\theta}$$

After cancelling out f_0 from Eq. (41.33) we obtain

$$g = -\frac{m v_y\,(v_x - \alpha y)}{\theta}\,\alpha\tau \tag{41.34}$$

Later on we shall need only the distribution at $y = 0$ to calculate viscous stresses on a wall. The symmetry of g is determined by the

product $v_x v_y$ or the product $\sin\vartheta\cos\vartheta\cos\varphi$ (for the case of a spherical function). With respect to the "magnetic quantum number", or the m (in the first volume it was denoted k) in the spherical function, there occurs a degeneracy [see Sec. 29]. Consequently, τ depends only upon l, which in the present case is equal to 2. The dependence of g upon the absolute value of v at $y = 0$ does not correspond to any eigenfunction of $\hat{\Lambda}_2$, at least for an arbitrary effective cross section $d\sigma/d\Omega_\chi$. Therefore no definite number τ_{2s} can be chosen: all values of τ_{2s} are involved in the exact answer.

But the largest value of τ_{2s} must be the decisive factor for the relaxation process, so to say the "bottleneck" on the way to equilibrium (see Exercise 2). At the given $l = 2$ the least eigenvalue of $\hat{\Lambda}_2$, which corresponds to the greatest relaxation time, must be taken. We denote it $(\tau_{20})^{-1}$. The 0 denotes that the least value of $(\tau_{20})^{-1}$ is yielded by the eigenfunction of the "ground state" of $\hat{\Lambda}_2$.

Let us now calculate the viscous stresses on a wall. They are given by the mean value of the momentum component transported in unit time on the wall in the y direction. Similarly to the calculation of the pressure on a wall in Section 2, we must now find the integral of $nf_0(1 + g)mv_xv_y$. The equilibrium distribution function makes no contribution. Therefore only the integral involving g remains:

$$p_{xy} = n\int_{-\infty}^{\infty} mv_x\,dv_x\int_0^{\infty} dv_y\int_{-\infty}^{\infty} dv_z\,f_0 g$$

$$= C\int_{-\infty}^{\infty} v_x^2\,dv_x\int_0^{\infty} v_y^2\,dv_y\int_{-\infty}^{\infty} dv_z\,\tau_{20}(v)\,e^{-mv^2/(2\theta)}$$

where $C = -(n\alpha m^2/\theta)\,[m/(2\pi\theta)]^{3/2}$. The lower limit in the integral over v_y can also be extended to $-\infty$, adding the factor 1/2. Initially it was taken only over the atoms travelling from the bulk to the wall. Furthermore, $v_x^2 = v^2\sin^2\vartheta\cos^2\varphi$, $v_y^2 = v^2\cos^2\vartheta$, and the integral over the total solid angle is equal to 4π multiplied by the mean square of $\cos^2\varphi$, that is, by 1/2, and the difference between the mean values of $\cos^2\vartheta$ and $\cos^4\vartheta$ over the solid angle $(1/3 - 1/5 = 2/15)$. As a result the integral over the angles yields $4\pi/5$ [Sec. 20].

The coefficient of visocsity η is the proportionality coefficient between $-\alpha = \partial\bar{v}_x/\partial y$ and p_{xy} (Sec. 17):

$$\eta = \frac{1}{15}\left(\frac{2}{\pi}\right)^{1/2}\frac{m^{7/2}}{\theta^{5/2}}\int_0^{\infty} v^6\tau_{20}(v)\,e^{-mv^2/(2\theta)}\,dv \tag{41.35}$$

It can be seen from Eqs. (41.26) and (41.27) that viscosity does not depend upon density, because the time τ is inversely propor-

tional to n, which cancels out in (41.35). This was theoretically discovered by Maxwell, who initially regarded it as a paradox.

If the interatomic force is a power function of the distance between the atoms, then the effective cross section is inversely proportional to $v^{4/n}$ (where n is the exponent in the potential energy of interaction: $U = A/r^n$). Then the relaxation time is proportional to $v^{-1+4/n}$, and the integral (43.25) over the velocities depends upon the temperature according to the law $\theta^{(1/2)(6+4/n)} = \theta^{3+2/n}$. Consequently $\eta \propto \theta^{3+2/n-5/2} = \theta^{1/2+2/n}$. A comparison with experiment makes it possible to determine the effective value of n (not to be confused with density). At $n = 4$, the time τ does not depend on velocity, and $\eta = n\tau\theta$.

Heat conductivity of a gas is determined analogously. But since heat flux is a vector, the Legendre polynomial is now P_1 (corresponding to symmetry of the problem) and, of course, the relaxation time is different (see Exercise 4).

To find the exact expression for the diffusion coefficient we must investigate a mixture of two gases, nonhomogeneous in concentration but with the same pressure at all points (otherwise a composite hydrodynamic gas flow would develop). It follows from the strict transport equation of the type (41.21) that the diffusion coefficient is a slowly varying function of the relative concentration of the components.

This can be qualitatively explained in the following way. Imagine a mixture of a light and heavy gas. Diffusion represents a relative displacement of one component in the other. The lighter component always makes the greater contribution since it is the more mobile (regardless of whether there is more or less of it). This is confirmed by calculations.

In diffusion in a medium with nonuniform temperature there is observed an effect predicted theoretically by D. Enskog. The temperature gradient itself causes a diffusion flux in a two-component mixture. This phenomenon is called *thermal diffusion.* Note that the elementary kinetic theory of gases, which is based only on the free-path concept, cannot give even the sign of the corresponding coefficient or the direction of the diffusion flux with respect to the temperature gradient.

Plasma Fluctuations. The transport equation involving a relaxation term in place of the integral term has the following form:

$$\frac{\partial f}{\partial t} + (\mathbf{v} \cdot \operatorname{grad} f) + \frac{d\mathbf{p}}{dt}\frac{\partial f}{\partial \mathbf{p}} + \frac{f_0 - f}{\tau} = 0 \qquad (41.36)$$

If the distribution function varies rapidly with time under the action of certain forces, for example, if a gas begins to vibrate with a period much smaller than the relaxation time, then the last term

in Eq. (41.36) may become negligibly small in comparison with the others. In other words, collisions between particles will be immaterial for the determination of f. The form of f in this case depends upon the motion of each particle in the field of all the other particles or in an external field.

Such a state occurs in a plasma subject to high-frequency vibration. Plasma, as is known, is an ionized gas comprising heavy positive ions and electrons (if ionization is not complete, plasma also contains neutral atoms). On the whole such a gas is neutral, that is, it contains equal numbers of ions and electrons, but in vibration local changes in the density of charges of both signs occur.

Suppose near some point the electron density has changed by a quantity n'. Then, in accordance with [16.1], an electric field appears and

$$\operatorname{div} \mathbf{E} = 4\pi e n'$$

This field acts upon the electrons in the usual way, that is

$$m \frac{d\mathbf{v}}{dt} = e\mathbf{E}$$

Assuming the density variations to be small, we can replace the total derivative of the velocity by the partial derivative, as was done in (16.5), where acoustic vibrations were considered. Take the divergence of both parts of the latter equation:

$$m \frac{\partial}{\partial t} \operatorname{div} \mathbf{v} = e \operatorname{div} \mathbf{E} = 4\pi n e^2$$

From the approximate form of the continuity equation (16.4), $\operatorname{div} \mathbf{v} = -(1/n_0)(\partial n'/\partial t)$, with n_0 the mean electron density. This yields an equation for n':

$$\frac{\partial^2 n'}{\partial t^2} = -\frac{4\pi n_0 e^2}{m} n'$$

Hence the electron density oscillates with a frequency

$$\omega_0 = \left(\frac{4\pi n_0 e^2}{m}\right)^{1/2} \qquad (41.37)$$

which is called the *Langmuir (plasma) frequency*. The corresponding oscillation frequency varies in inverse proportion to the square root of the density, whereas the relaxation time is inversely proportional to the density itself. It follows from this that in sufficiently rarified plasma the relaxation component always becomes small enough in comparison with other terms of the transport equation.

Let us now determine the mean energy of plasma oscillations. The mean kinetic energy of an individual electron is

$$\frac{1}{2} m\overline{v^2} = \frac{m}{2} \frac{e^2 \mathbf{E}_0^2 \overline{\cos^2 \omega_0 t}}{m^2 \omega_0^2} = \frac{e^2 \mathbf{E}_0^2}{4m\omega_0^2}$$

where $\mathbf{E}_0$ is the maximum value of the field. Multiplying by the electron density, we find:

$$\frac{1}{2} n_0 m\overline{v^2} = \frac{\mathbf{E}_0^2}{16\pi} = \frac{\overline{\mathbf{E}^2}}{8\pi} \tag{41.38}$$

In other words, the mean energy of the oscillations of electrons is equal to the mean energy of the field.

Now write the transport equation for plasma without the relaxation term:

$$\frac{\partial f}{\partial t} + (\mathbf{v} \cdot \mathbf{grad}\, f) + \frac{e\mathbf{E}}{m} \frac{\partial f}{\partial \mathbf{v}} = 0 \tag{41.39}$$

To this we must add the equation for $\mathbf{E}$:

$$\operatorname{div} \mathbf{E} = 4\pi e \int f \, d\tau_{\mathbf{v}} \tag{41.40}$$

where the function f is normalized to n'. This equation was developed by A.A. Vlasov.

Since (41.40) involves a derivative with respect to position, it means that not only waves of the type $\mathbf{E} = \mathbf{E}_0 \cos \omega_0 t$, that is, spatially homogeneous waves, can exist in plasma but also acoustic-type travelling longitudinal waves $E_x = E \cos(kx - \omega t)$. At small k, or long wavelengths, their frequency ω approaches the Langmuir plasma frequency ω_0.

Landau Damping. L. D. Landau discovered that in plasma travelling waves are to some extent attenuated owing to energy transfer to individual electrons. This attenuation is not accompanied by any increase in entropy since it is not a result of collisions, that is, it is not associated with relaxation processes. The energy received by individual electrons from the collective, hydrodynamic motion of plasma may then be returned by them to the field and to the motion of the plasma as a whole. In true damping, that is, conversion of wave energy into heat, this is impossible.

Landau obtained his result directly from the Vlasov equations. The same can be deduced from equations of mechanics by more lengthy but more visual and elementary computations. We shall adopt the latter method.

Let a longitudinal wave be travelling along the x axis. Then the motion of an electron is subject to the equation

$$m \frac{d^2x}{dt^2} = eE_0 \cos(kx - \omega t) \tag{41.41}$$

We put the initial conditions at $t = 0$ as follows: $x = x_0$, $dx/dt = v_0$, and introduce a new unknown $\xi \equiv kx - \omega t$. The initial conditions then are $\xi_0 = kx_0$, $\dot{\xi}_0 = kv_0 - \omega$. It is subject to the

equation

$$m\frac{d^2\xi}{dt^2}=eE_0k\cos\xi \tag{41.42}$$

which is easily solved explicitly. Multiply both parts by $d\xi = dt(d\xi/dt)$ and take the first integral:

$$\frac{m}{2}\left(\frac{d\xi}{dt}\right)^2=eE_0k\sin\xi+C$$

The initial conditions yield

$$C=\frac{mv_0^2}{2}-eE_0k\sin kx_0$$

Then

$$\left(\frac{d\xi}{dt}\right)^2=(kv_0-\omega)^2+\frac{2eE_0k}{m}(\sin\xi-\sin kx_0)$$

Extracting the square root, separating the variables, and integrating once again, we obtain

$$t=\int_{kx_0}^{kx_0-\omega t}\frac{d\xi}{\left[(kv_0-\omega)^2+\frac{2eE_0k}{m}(\sin\xi-\sin kx_0)\right]^{1/2}}$$

$$=\int_{kx_0}^{kx_0-\omega t}\frac{d\xi}{R} \tag{41.43}$$

Since we have to determine dx/dt, or the velocity of an electron, we differentiate the obtained expression with respect to x. This yields

$$\frac{dt}{dx}=\frac{1}{R}\left(k-\omega\frac{dt}{dx}\right)$$

Solving with respect to $\frac{dx}{dt}$, we find

$$\frac{dx}{dt}=\frac{1}{k}(R+\omega) \tag{41.44}$$

The latter equation shows that the root should be taken with the same sign as that of the quantity $kv_0-\omega$ (to satisfy the initial condition for the velocity). From (41.44) we must determine the kinetic energy of the electron (more precisely, the term $mv_x^2/2$) and average it over the initial coordinate and velocity of the electron.

We first find the average over the coordinate x_0. Assuming the electric field to be weak, we expand the root in a series up to the term quadratic with respect to the field, that is, for $(1+\alpha)^{1/2}$ we substitute $1+\alpha/2-\alpha^2/8$. This yields

$$\frac{m}{2}\left(\frac{dx}{dt}\right)^2=\frac{mv_0^2}{2}-\frac{\omega e^2E_0^2}{2m^2}\,\frac{[\sin(kx-\omega t)-\sin kx_0]^2}{(kv_0-\omega)^3}$$

$$+\text{terms linear in } E_0$$

Into this we must substitute $x = v_0 t + x_0$. Then in averaging over x_0 the linear terms vanish. Next we transform the term in brackets:

$$\sin(kx - \omega t) - \sin kx_0 = \sin[kv_0 + t(kx_0 - \omega)] - \sin kx_0$$
$$= 2\cos\left[kx_0 + \frac{t(kv_0 - \omega)}{2}\right]$$
$$\times \sin\frac{t(kv_0 - \omega)}{2}$$

In averaging over x_0 the square of the cosine in the first term yields 1/2. Therefore the mean value of $mv_x^2/2$ for an electron with an initial velocity v_0 is

$$\overline{\frac{mv_x^2}{2}} = \frac{mv_0^2}{2} - \frac{\omega e^2 E_0^2}{m} \frac{\sin^2 \frac{t}{2}(kv_0 - \omega)}{(kv_0 - \omega)^3} \tag{41.45}$$

Now we carry out the averaging over the initial velocity. The integrals of the Maxwell distribution over the other velocity components yield unity. There remains the factor

$$f_0(v_0) = \left(\frac{m}{2\pi\theta}\right)^{1/2} e^{-mv_0^2/(2\theta)} \tag{41.46}$$

We must calculate the following integral:

$$\overline{\Delta\frac{mv_x^2}{2}}^{x_0,\, v_0} = -\frac{\omega e^2 E_0^2}{m} \int_{-\infty}^{\infty} \frac{\sin^2 \frac{t}{2}(kv_0 - \omega)}{(kv_0 - \omega)^3} f(v_0)\, dv_0 \tag{41.47}$$

where the superscripts x_0 and v_0 indicate the quantities over which the averaging was carried out. We obtain an integral which can be formally assumed to diverge at $v_0 = \omega/k$. Actually this is connected with the expansion of the square root in a series: without the expansion the result would be single valued and finite. That means that the integral (41.47) must be taken in the sense of its principal value (Sec. 36), which is also finite. Indeed, first exclude the section $\omega/k - \varepsilon \leqslant v_0 < \omega/k + \varepsilon$ from the integration interval. Tend ε to zero in the exact expression: this does not introduce any infinite terms. Going over to the same limit in the approximate integral (41.47) yields the principal value.

To find it, expand $f_0(v_0)$ in a series in the vicinity of $v_0 = \omega/k$ up to the first power:

$$f_0(v_0) = f_0\left(\frac{\omega}{k}\right) + \left(v_0 - \frac{\omega}{k}\right) f_0'\left(\frac{\omega}{k}\right)$$

The integral with the zero term of the expansion is taken with respect to an odd function, and it yields a principal value equal to zero. We encountered the integral of the first term in [32.39]. Making

the substitutions $t(kv_0 - \omega)/2 = \xi$ and taking advantage of the fact that

$$\int_{-\infty}^{\infty} \frac{\sin^2 \zeta}{\zeta^2}\, d\zeta = \pi$$

(see [32.40]), we obtain

$$\Delta \frac{m}{2} \overline{\left(\frac{dx}{dt}\right)^2}^{x_0,\, v_0} = \frac{\pi}{2} \frac{\omega e^2 E_0^2 t}{mk^2} f_0' \left(\frac{\omega}{k}\right) \tag{41.48}$$

To simplify this formula we take advantage of the energy conservation law: the total energy of the electrons, field and plasma oscillations must be conserved, or, from (41.38), the relationship

$$\frac{d}{dt} \frac{E_0^2}{8\pi} = -\frac{d}{dt} \Delta \frac{n_0 m}{2} \overline{\left(\frac{dx}{dt}\right)^2}^{x_0,\, v_0} \tag{41.49}$$

should hold. If γ is the damping factor of the field amplitude, then

$$\frac{d}{dt} \frac{E_0^2}{8\pi} = -2\gamma \frac{E_0^2}{8\pi} = \frac{\pi}{2} \frac{n_0 \omega e^2 E_0^2}{mk^2} f_0' \left(\frac{\omega}{k}\right) \tag{41.50}$$

whence

$$\gamma = 2\pi^2 \frac{n_0 e^2 \omega}{mk^2} f_0' \left(\frac{\omega}{k}\right)$$

Assuming the wave vector to be small (this requirement will be clarified in a moment), we substitute ω_0 for ω. Then γ reduces to

$$\gamma = -\frac{\pi}{2} \frac{\omega_0^3}{k^2} f_0' \left(\frac{\omega_0}{k}\right) \tag{41.51}$$

Introducing the notation $\varkappa = (4\pi n_0 e^2/\theta)^{1/2}$, we find that substitution into (41.46) yields the result obtained by Landau:

$$\gamma = \left(\frac{\pi}{8}\right)^{1/2} \left(\frac{\varkappa}{k}\right)^3 \exp\left[\frac{1}{2}\left(\frac{\varkappa}{k}\right)^2\right] \tag{41.52}$$

From this we see why we can call k a small quantity: the inequality $k \ll \varkappa$ must be satisfied or, in other words, the wavelength must be long in comparison with $\varkappa$.

The physical origin of γ is apparent from the formulas. If $f_0\,(v_0)$ is a decreasing function of velocity, then close to $v_0 = \omega/k$ there is a slight surplus of the slower electrons over the faster ones. Electrons travelling at speeds close to the phase velocity of the wave are swept on by its motion, as it were falling in step with it. But since there are more slower electrons, as a result they take energy from the wave. That is the cause of damping.

EXERCISES

1. Particles capable of multiplying diffuse in a spherical volume of radius a. They may be, for example, neutrons in uranium enriched with the isotope U_{235}, or active centres of a branching chemical chain reaction. This is taken into account in the diffusion equation by adding the term αn to $D\Delta n$ (where $\alpha > 0$). Then under homogeneous conditions ($\Delta n = 0$) $n = n_0 e^{\lambda t}$. Actually, the particles that reach the surface of the sphere drop out of the process (neutrons fly out, active centres of a chemical reaction recombine). This is approximately described by the boundary condition $n(a) = 0$. Determine the value of the radius at which the number of particles begins to increase exponentially with time (the *critical size*).

Solution. The solution possesses central symmetry and must therefore be sought in the form $n = f(r)\, e^{-\lambda t}/r$. At $\lambda > 0$ the concentration damps out with time. The function $f(r)$ satisfies the equation

$$D\frac{d^2f}{dr^2} = -(\alpha+\lambda) f$$

with the boundary condition $f(0) = f(a) = 0$. Since f is essentially a positive quantity, we select a solution without nodes:

$$f = \sin\left(\frac{\alpha+\lambda}{D}\, r^2\right)^{1/2}$$

From this follows the equation for λ:

$$a\left(\frac{\alpha+\lambda}{D}\right)^{1/2} = \pi$$

The quantity λ vanishes when the radius is

$$a = \pi\left(\frac{D}{\alpha}\right)^{1/2}$$

At large values of the radius, as was shown by V. G. Bursian and V. S. Sorokin, the number of particles increases exponentially.

2. A monatomic gas is disturbed from equilibrium by an external disturbance described by a function of the form $h(v)\, Y_l^m$ satisfying the equation

$$\frac{\partial g}{\partial t} + \hat{L} g = h(v)\, Y_l^m$$

The disturbance started at the initial time $t = 0$. The operator $\hat{L}$ is defined by formulas (41.26) and (41.27). Determine the asymptotic form of g at $t = \infty$.

Solution. We separate the angular part of g by substitution of (41.28):

$$g(v) = G_l(v)\, Y_l^m$$

This yields for $G_l(v)$ the equation:

$$\frac{\partial G_l}{\partial t} + \hat{\Lambda}_l G_l = h(v)$$

Expand $h(v)$ in terms of the complete set of eigenfunctions of the operator $\hat{\Lambda}_l$: $h=\sum c_s G_{ls}$. We seek the solution in the form $G_l = \sum b_s(t)\, G_{ls}$. Then each "amplitude" b_s satisfies the equation

$$\frac{db_s}{dt}+\frac{b_s}{\tau_{ls}}=c_s$$

Since at time zero the disturbance is zero, we obtain the solution

$$b_s(t)=\tau_{ls}c_s\left(1-e^{-t/\tau_{ls}}\right)$$

Asymptotically it has the form

$$b_s=\tau_{ls}c_s$$

If one of the relaxation times, for instance, τ_{l0}, appreciably exceeds all the others, only the first term need be retained:

$$G_l(t \longrightarrow \infty) \approx \tau_{l0}h$$

which corresponds to the adopted approximation.

3. A monatomic neutral gas contains individual ionized atoms (of the same or some other gas). The gas is placed in an external electric field. Determine the mobility of the ions.

Solution. As was shown, the force between an ion and an atom decreases according to the Maxwellian law, so that the probability of their colliding does not depend upon their relative velocity. At low ion concentration the distribution function f_0 of the atoms can be assumed independent of the presence of ions. Then for the ion distribution function we obtain the transport equation

$$\frac{\partial f}{\partial t}+\frac{e\mathbf{E}}{m}\frac{\partial f}{\partial \mathbf{v}}=-n\int\int |\mathbf{v}-\mathbf{v}'|\,d\sigma\, d\tau_{\mathbf{v}'} \times [f(\mathbf{v})\,f_0(\mathbf{v}')-f(\mathbf{v}_1)\,f_0(\mathbf{v}_1')]$$

In steady-state conditions $\partial f/\partial t = 0$. The problem can be solved without finding the ion distribution function $f(\mathbf{v})$.

Let the field be directed along the x axis. Multiply both parts of the equation by v_x and integrate with respect to $d\tau_{\mathbf{v}}$. On the left-hand side we have

$$\frac{eE_x}{m}\int v_x\frac{\partial f}{\partial v_x}\,d\tau_{\mathbf{v}}=-\frac{eE_x}{m}\int f\,d\tau_{\mathbf{v}}=-\frac{eE_x}{m}$$

Now calculate the integral of the product of v_x and the right-hand side of the equation. In the first term we obtain

$$-n\int f v_x\,d\tau_{\mathbf{v}} \times \int f_0(\mathbf{v}')\,d\tau_{\mathbf{v}'}\int |\mathbf{v}-\mathbf{v}'|\,d\sigma$$

The first factor is equal to $\overline{v_x}$, the second to 1 (from the normalization condition), and the integrand of the third factor is independent of the velocity (from the basic property of collisions of particles obeying the Maxwellian interaction law).

In the second term of the integrand on the right it is convenient to take $\mathbf{v}_1$ and $\mathbf{v}_1'$ as the integration variables. The Jacobian of the transformation from $\mathbf{v}$, $\mathbf{v}'$ to $\mathbf{v}_1$, $\mathbf{v}_1'$ is unity, which can be verified by direct computation. Therefore, for integration $\mathbf{v}$ must be expressed in terms of $\mathbf{v}_1$, $\mathbf{v}_1'$ and the angle of turn of the relative velocity in the centre-of-mass frame of reference.

The transformation is carried out in the familiar way. If a particle of mass m and velocity $\mathbf{v}$ collides with a particle of mass M and velocity $\mathbf{v}'$, then in the centre-of-mass reference frame their velocities prior to the collision are

$$\mathbf{v}_{\text{c. m}}=\frac{M(\mathbf{v}-\mathbf{v}')}{M+m}, \qquad \mathbf{v}'_{\text{c. m}}=-\frac{m(\mathbf{v}-\mathbf{v}')}{M+m}$$

Correspondingly, after the collision

$$\mathbf{v}_{1\text{c. m}}=\frac{AM(\mathbf{v}-\mathbf{v}')}{M+m}, \qquad \mathbf{v}'_{1\text{c. m}}=-\frac{Am(\mathbf{v}-\mathbf{v}')}{M+m}$$

where A is an abbreviated symbol of the matrix of the cosines of the angle of turn of the vector (see Exercise 3, [Sec. 4]). Returning to the initial reference system, we obtain

$$\mathbf{v}_1=\frac{(AM+m)\,\mathbf{v}+M(1-A)\,\mathbf{v}'}{M+m},$$

$$\mathbf{v}_1'=\frac{m(1-A)\,\mathbf{v}+(Am+M)\,\mathbf{v}'}{M+m}$$

Cancelling out $\mathbf{v}'$, we find the required expression for $\mathbf{v}$:

$$\mathbf{v}=\frac{(m+A^{-1}M)\,\mathbf{v}_1}{M+m}+\frac{M(1-A^{-1})\,\mathbf{v}_1'}{M+m}$$

where A^{-1} is the inverse of A. Since $f_0(\mathbf{v}')$ is the equilibrium Maxwellian function, in integrating over the directions of $\mathbf{v}'$ the second term becomes zero. All that remains of the operator A^{-1} is the element $(A^{-1})_{xx}=\cos\chi$, because in integrating along the azimuth of the rotated velocity vector, $(A^{-1})_{xy}$ and $(A^{-1})_{xz}$ become zero.

Therefore

$$\frac{eE_x}{m}=n\overline{v_x}\int|\mathbf{v}-\mathbf{v}'|\,d\sigma\left(1-\frac{m+M\cos\chi}{m+M}\right)$$

It is significant that in the present case the product $|\mathbf{v}-\mathbf{v}'|\,d\sigma$ depends on χ. The obtained integral can be found together with its numerical coefficient. We have thus found the relationship between an applied field and the directed component of the electron velocity (the *drift velocity*). The result agrees well with experiment.

4. Determine the heat conductivity of an ideal monatomic gas.

Solution. Supposing that as a whole the gas must be at rest, in the computations the pressure must be assumed constant. Then only energy transfer occurs; there is no transfer of mass. Assuming the temperature to vary with

the coordinate according to a linear law, we represent it as $\theta = \theta_0 + \theta'_x$ (where θ' is the temperature gradient). We seek the distribution function in the form

$$f = \left[\frac{m^3\theta_0^2}{(2\pi)^3(\theta_0+\theta'x)^5}\right]^{1/2} \exp\left\{-\frac{mv^2}{2(\theta_0+\theta'x)}\right\}[1+g(\mathbf{v}, x)]$$

In integration with respect to the velocity, g makes no contribution. Therefore

$$\int f\, d\tau_{\mathbf{v}} = \frac{\theta_0}{\theta_0+\theta'x}$$

Consequently the total number of atoms $n(x)$ in unit volume is equal to $n\theta_0(\theta_0 + \theta'x)^{-1}$, whence $p = n(x)\,\theta(x) = n\theta_0 =$ constant. Thus the constancy of pressure is assured.

Substituting the function f into the relaxation equation, we find

$$v_x\theta'\left(\frac{mv^2}{2(\theta_0+\theta'x)^2} - \frac{5}{2}\frac{1}{\theta_0+\theta'x}\right) = \frac{g}{\tau}$$

The meaning of the relaxation time in this equation will be clarified later. The correction to the distribution function at $x = 0$ is

$$g = -\frac{\theta'\tau v_x}{\theta_0}\left(\frac{mv^2}{2\theta_0} - \frac{5}{2}\right)$$

We define the heat flux q_x as the mean value of $(mv^2/2)v_x$ (substituting θ for θ_0):

$$q_x = -\frac{n\theta'\tau}{\theta}\frac{m}{(2\pi\theta)^{3/2}}\int d\tau_{\mathbf{v}}\, e^{-mv^2/(2\theta)}\, v_x^2\left(\frac{mv^2}{2} - \frac{5}{2}\right)\frac{mv^2}{2}\,\tau(v)$$

Integration over the angles yields $4\pi/3$; whence

$$q_x = -\frac{4\pi n\theta'}{3\theta}\frac{m}{(2\pi\theta)^{3/2}}\int_0^\infty v^4\, dv\,\frac{mv^2}{2}\left(\frac{mv^2}{2\theta} - \frac{5}{2}\right) e^{-mv^2/(2\theta)}\,\tau(v)$$

If $\tau(v)$ does not depend on the velocity, computations yield (see Exercise 3, Section 1)

$$q_x = -\frac{5}{2}\frac{\theta n\tau}{m}\theta'$$

Here the factor of $-\theta'$ is the heat conductivity of the gas. Note that it does not involve the relaxation time on which η depends, because here the symmetry of g is that of a vector, that is, a spherical function of order 1, and not of the tensor $v_x v_y$. The greatest relaxation time in a heat conductivity problem is characterized by the numbers $l = 1$, $s = 1$. To the value $s = 0$ at $l = 1$ corresponds $g = \mathbf{a}\mathbf{v}$ (where $\mathbf{a}$ is a constant vector). But such a function reduces the collision integral to zero by virtue of the momentum conservation law in collisions. Therefore $\tau_{10} = \infty$, which would yield infinite heat conductivity.

The function $\mathbf{a}\mathbf{v}$ is orthogonal, with a factor $v^2e^{-mv^2/2\theta}$, to $(mv^2/2\theta - 5/2)\mathbf{v}$ and is not involved in the expansion of the right-hand side of the equation

considered in Exercise 2. Therefore no infinite relaxation time τ_{10} is involved in the heat conductivity problem. The quantity q_x is expressed in terms of τ_{11}.

5. Find another solution of Eq. (41.27) corresponding to infinite relaxation time.

42

ELECTRONS IN CRYSTALS

The kinetics of electron phenomena in crystalline bodies originated in the works of Lorentz and Drude soon after the discovery of electrons as elementary particles. As was pointed out in Section 6, the main difficulty in the classical electron theory of metals was connected with the anomalously low specific heat of electrons in metals. This difficulty was removed by A. Sommerfeld, who applied Fermi statistics to them. But the real development of the electron theory of metals, and in general the quantum theory of crystalline state, began with the works of Felix Bloch carried out in 1930. Bloch established the main features of electron behaviour in crystalline lattices.

Motion of an Electron in a Periodic Field. The motion of an electron in a field possessing spatial periodicity could have been examined in quantum mechanics, but there the question would have remained without any applications. To establish the basic features of the problem we shall commence with a one-dimensional model.

Let the potential energy of an electron depend periodically on the x coordinate:

$$U(x+na)=U(x) \tag{42.1}$$

where n is an integer of any sign, and $-\infty \leqslant x \leqslant \infty$. Hence a is the spatial period of $U(x)$. Substitute the relationship (42.1) into the Schrödinger equation to get

$$-\frac{h^2}{2m}\frac{d^2\psi}{dx^2}+U(x+na)\psi=-\frac{h^2}{2m}\frac{d^2\psi}{dx^2}+U(x)\psi=E\psi \tag{42.2}$$

which does not change its form in such a substitution. Therefore, in the substitution of $x+a$ for x the wave function $\psi(x)$ can be multiplied only by a constant C whose modulus is unity:

$$\psi(x+a)=C\psi(x) \tag{42.3}$$

Indeed, $\psi(x + na) = C^n\psi(x)$. It can be seen from this that at $|C| \gtrless 1$ the function $\psi(x + na)$ tends to infinity at large n, positive or negative depending on $|C|$.

Any number of modulus unity can be represented in the form

$$C = e^{2\pi i\xi} \tag{42.4}$$

where ξ is a real quantity. It can be seen from this that the wave function satisfies the equation

$$\psi(x + na) = e^{2\pi i\xi n}\psi(x) \tag{42.5}$$

which holds for any function of the form

$$\psi(x) = e^{2\pi i\xi x/a}u_\xi(x) \tag{42.6}$$

Here $u_\xi(x) = u_\xi(x + na)$, that is, the function $u_\xi(x)$ has a period equal to the period of the lattice a. Substituting $x + na$ for x in (42.6), we find that the requirement (42.5) is satisfied.

The first factor in (42.6) is very much like the wave function of a free electron, $e^{ipx/h}$. To the momentum p corresponds the quantity

$$\frac{2\pi h\xi}{a} \equiv k \tag{42.7}$$

which is called the *quasi-momentum*.

Thanks to the periodicity of the potential field the motion of an electron in it is very like free motion. The distinction of the corresponding wave function consists in a modulating oscillating factor. This difference affects the dependence of the energy on the quasi-momentum, which can be found from the Schrödinger equation by substituting the wave function $\psi(x) = e^{ikx/h}u_k(x)$ into it:

$$-\frac{h^2}{2m}\frac{d^2u_k}{dx^2} - \frac{2ik}{mh}\frac{du_k}{dx} + \frac{k^2}{2m}u_k + U(x)u_k = Eu_k \tag{42.8}$$

Energy of an Electron in a One-dimensional Periodic Field. Quasi-momentum is involved as a parameter in Eq. (42.8). Therefore the energy eigenvalue depends upon k. The form of the dependence is determined from the periodicity condition for u_k. Thus the energy and quasi-momentum of an electron are in the same state as the energy and momentum of a free electron. But the dependence of the energy on k has a very complex form. The purpose of this section is mainly to explain certain basic features of the behaviour of $E(k)$ in one-dimensional and three-dimensional periodic structures.

To begin with, energy is an even function of quasi-momentum: $E(k) = E(-k)$. Indeed, a transition from ξ to $-\xi$ or, what is the same, from k to $-k$, corresponds to a substitution of the wave function by a complex conjugate one. But the transition does not affect the energy eigenvalue, since it means the substitution of $-t$ for t. Quantum mechanical equations are symmetrical with respect to a re-

versal of time (in the presence of an external magnetic field its sign also has to be changed).

We shall now find the general form of the curve representing the dependence of E on k. For this note that Eq. (42.4) defines the real quantity ξ up to an integral term, positive or negative. It is therefore convenient to agree to refer ξ always to the same interval $-1/2 \leqslant \xi \leqslant +1/2$. If ξ lies outside this interval, an integer can always be added that will bring it back between 1/2 and $-1/2$. We shall call this integer n.

Equation (42.8) involves the number ξ itself, or k. Hence each value of n corresponds to different energies and different functions u_k. Assuming, however, that ξ lies between $-1/2$ and 1/2, we can provide the energy with an additional subscript n:

$$E = E_n(k)$$

Here it is sufficient to take only positive n since energy is an even function of quasi-momentum.

It is convenient to introduce the second index n instead of the continuously varying k for the following reason. At $\xi \pm 1/2$ Eq. (42.3) yields

$$\psi(x+a) = e^{\pm i\pi}\psi(x) = -\psi(x) = \psi(x-a)$$

These equations contain no imaginary numbers, therefore such a wave function does not correspond to the propagation of any wave [Sec. 22]. It describes a standing wave, which is constructed in the following way.

Equation (42.8) involves i. A solution of the standing-wave type is obtained in the form of linear combinations $u_k + u_{-k}$ and $i^{-1}(u_k - u_{-k})$ at $k = \pi h/a$, just as $\cos x$ and $\sin x$ can be obtained from e^{ix}. But $u_k + u_{-k}$ and $u_k - u_{-k}$ are different wave functions to which correspond different energy values. Furthermore, since $E(k)$ is an even function of k, these values occur at the ends of the interval $-\pi h/a \leqslant k \leqslant +\pi h/a$. The same occurs every time when $k = \pi hn/a$, $\xi = n/2$, that is, at all integral and half-integral ξ.

But if a function has two values for the same argument k and is single-valued at neighbouring points, this reduces simply to a discontinuity. In other words, within the interval with every value of n the function $E_n(k)$ is continuous, while in going over to the next interval it experiences a finite discontinuity. The curve consists of separate smooth sections, which are conveniently referred always to the interval $-1/2 \leqslant \xi \leqslant +1/2$, and are drawn one over the other as shown in Figure 54.

The hatched areas indicate the energy intervals which cannot contain the electron's energy in the given periodic field. Each section of permitted energy values is called the *energy band* of the given number n. Between them lie forbidden energy bands.

The origin of such an energy spectrum can be explained as follows. Let period a tend to infinity. Then the state of the electron in the periodic field is replaced by the aggregate states around each separate attraction centre. Such centres model the field of atomic cores in a crystal lattice. Individual states here correspond to allowed energy bands, which are contracted into levels. When the period decreases, the electron's action on all other centres comes into play (except

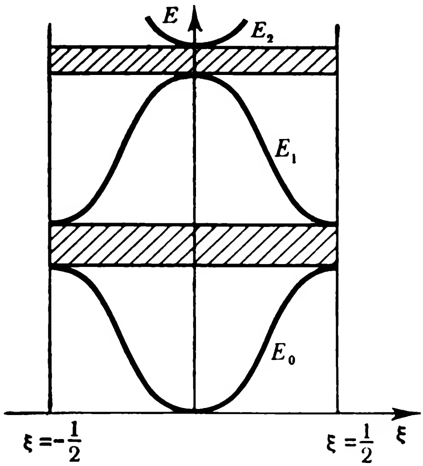


Figure 54

the centre with which it is associated). This disturbance removes the degeneration of the level consisting in that one electron can be related to all centres. The level splits according to the number of separate centres. But since in a real crystal the number of centres is enormous, what occurs is not a splitting into discrete levels but a broadening of the degenerate level, so that a band forms. The weaker bound states, to which higher levels correspond, broaden more.

The wave functions (42.6) are essentially the true eigenfunctions of the degenerate problem defined by the quantum number k and the band number n [32.20].

The eigenfunctions of (42.2) comprise a complete set. In principle they can be used to construct any wave packet describing an electron's motion in a periodic field $U(x)$. If we restrict ourselves to states belonging to the same band, which corresponds to $\Delta k = 2\pi h/a$, it follows from the uncertainty relations [22.4] that $\Delta x = a$, that is, localization of an electron is possible within one cell (one period).

As we know from quantum mechanics (see [21.7]), the velocity of propagation of a wave packet is the velocity of the electron. The

relationship $v = \partial E/\partial p$ in a periodic field is changed to

$$\left(\frac{dx}{dt}\right)_n = v_n = \frac{\partial E_n}{\partial k} \tag{42.9}$$

But here the coordinate is determined up to one period. The subscript n indicates that the electron belongs to the nth energy band.

Formula (42.9) shows that v is a function of k, that is, that the velocity is a constant of motion. It exists in the same state as the energy, as in the case of a free electron. This is an extremely important aspect of the resemblance between an electron in a periodic field and a free electron. In such a field electrons move without scattering on periodically located centres.

If the periodic function of potential energy is stated exactly, the form of the energy spectrum similar to that shown in Figure 54 is in principle defined completely. An example of such definition is offered in Exercise 1.

Weakly Bound Electrons. Here we shall determine the spectrum approximately, assuming that in the zero approximation the electron is free and described by the wave function $e^{ikx/h}$ (where k is the ordinary momentum). The potential energy will be treated as a perturbation. Sometimes this is also justified in the quantitative sense. If a crystal is made up of complex atoms, the charge of the nucleus is screened by electrons of the atomic core. If the band characterizing the state in the lattice is correlated to the state in the atom, the outer electron, whose motion is considered, may have a large quantum number. At high quantum numbers the wave function rapidly oscillates within the limits of the atomic core. As is known [Sec. 33], each unit in the quantum number shows how many nodal surfaces the wave function possesses. Therefore, the mean value of the potential acting on the outer electron corresponds to a matrix element involving rapidly oscillating wave functions. As a consequence we find that the periodic part of the outer electron's wave function falls in the domain in which the electron interacts with the periodic field of the lattice. Such an effectively weakened potential is called a *pseudopotential.*

The kinetic energy of an electron in a crystal was estimated in Section 6. It is of the order of several volts. The pseudopotential is of the order of 0.1-0.2 V. Therefore, if the pseudopotential is small, the correction to the electron's energy introduced by the lattice can be determined from the perturbation theory.

The first-approximation correction is equal to the mean value of the potential over the volume. It does not involve the periodic part of the potential at all: by due energy calibration its mean value is selected equal to the work function of an electron in escaping from the crystal, computed from the bottom of the energy band. If the

band corresponds to the state of an individual atom with ionization energy E_n, which in the crystal becomes the allowed band of energy values, the electron's energy in the band can be treated as kinetic, taking the least energy of the band for zero.

Let us now calculate the correction to the energy due to the periodic part of the potential. If it is assumed small, the electron must in the first approximation be regarded as free (taking into account its energy shift due to its bond with the lattice as a whole). This does not affect the form of the wave function, which, as in the case of a free electron, is $e^{ikx/h}$. The energy of such an electron is equal to $k^2/(2m) - |E_0|$ (where $|E_0|$ is the binding energy; $E_0 = \langle U \rangle$).

From [32.15], in the second approximation the correction to the energy for a continuous spectrum must have the form

$$E_k^{(2)} = -\int \frac{|U_{kk'}|^2}{E_k - E_{k'}} dk' \tag{42.10}$$

Here $U_{kk'}$ is a matrix element of the periodic part of the potential (V is the volume of the crystal, introduced for normalization):

$$U_{kk'} = \frac{1}{V} \int e^{i(k-k')x/h} U(x)\, dx \tag{42.11}$$

The quantity $U(x)$ can be expanded in a Fourier series:

$$U(x) = \mathrm{Re}\left[\sum_{n=1}^{\infty} U_n e^{2\pi inx/a}\right] \tag{42.12}$$

The value $n = 0$ is excluded, since by definition the mean value is E_0. Then, as can be seen from (42.11), the matrix element differs from zero only at $|k - k'| = 2\pi nh/a$, and in this case is equal to the Fourier coefficient in the expansion of the potential.

The denominator of expression (42.11) vanishes in two cases: at $k = k'$ and at $k = -k'$. In the former case the numerator also becomes zero because there is no Fourier coefficient with $n = 0$. In the latter case in the numerator we have the integral

$$U_{k\,-k} = \frac{1}{V} \int_{-\infty}^{\infty} e^{2ikx/h} U(x)\, dx \tag{42.13}$$

It is not zero in those cases when $k = \pi hn/a$, that is, when k is located at the edge of the allowed energy band. Therefore Eq. (42.10), in which the correction to the energy is computed as a small perturbation, is inapplicable in such form. The electron's wave function is strongly disturbed, while the energy, though it varies relatively weakly, does not appear as an expansion in powers of the perturbation.

The energy of the unperturbed problem is always degenerate: $E(k) = E(-k)$. But at $k \neq \pi hn/a$ the matrix element $U_{k\ -k}$ between these two states becomes zero, and the perturbation does not have a strong effect.

At $k = \pi hn/a$ the matrix element is, from (42.13), equal to the coefficient U_n of the Fourier expansion of the potential. As we know from [Sec. 32], the wave function must be taken in the form of a linear combination of both functions corresponding to the degenerate energy value:

$$\psi = c_k \psi_k + c_{-k} \psi_{-k}$$

It must be substituted into the wave function, multiplied from the left by ψ_k^* and ψ_{-k}^*, and integrated. Taking into account the orthogonality of wave functions, we obtain two linear homogeneous equations:

$$\left(\frac{k^2}{2m} - |E_{0n}|\right) c_k + U_n c_{-k} = E c_k$$

$$\left(\frac{k^2}{2m} - |E_{0n}|\right) c_{-k} + U_n^* c_k + E c_{-k}$$

For them to have a solution the system determinant must become zero:

$$\left(E - \frac{k^2}{2m} + |E_{0n}|\right)^2 - |U_n|^2 = 0$$

It can be seen from this that the energy of an electron in a lattice differs from the energy of a free electron by the quantity

$$E - \frac{k^2}{2m} = \pm |U_n| - |E_{0n}| \tag{42.14}$$

Thus, at the edge of the band a discontinuity of $2\,|U_n|$ occurs, that is, $2\,|U_n|$ is the width of the forbidden band in the considered approximation. Although we assume U small in comparison with the energy E, formula (42.14) should, nevertheless, not be viewed as a power series in a small quantity. As is known from calculus, close to points where a function undergoes a discontinuity it cannot be expanded in a power series because it is nondifferentiable. The behaviour of the energy in the neighbourhood of the discontinuity point is examined in Exercise 2.

In accordance with the sign in front of $|U_n|$ in (42.14), the ratio of the coefficients, c_k/c_{-k}, is ± 1. Thus, the wave function has the form of a standing wave, as it should be on the edge of a band.

The velocity of an electron on the edge of the band, $v = \partial E/\partial k$, is zero (see Exercise 2).

Let us now construct the energy spectrum of a weakly bound electron in a periodic field. We lay off the momentum on the abscissa

(Figure 55). Then, if on the ordinate we lay off only the part of the energy dependent upon the momentum, in the zero approximation we obtain a parabola $k^2/(2m)$, shown by the dashed line. The solid curve, or rather the solid segments, show the energy curve taking into account the discontinuities at $k = \pi hn/a$. At these points $\partial E/\partial k = 0$. Far away from the discontinuities the solid and dashed curves lie very close together.

To make a construction of the bands similar to the one in Figure 54, the segments of the solid curve must be shifted to the section

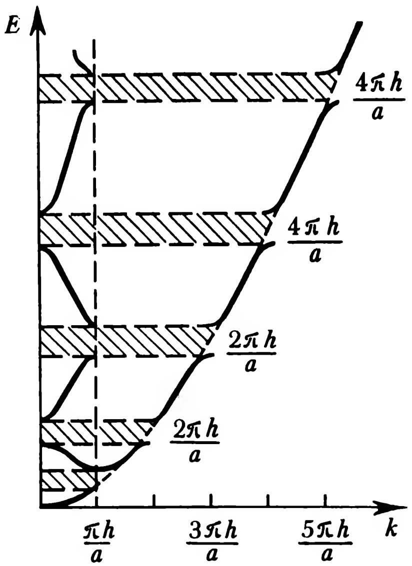


Figure 55

$-\pi h/a \leqslant k \leqslant +\pi h/a$, so that the points corresponding to the even discontinuities be on the ordinate, and those corresponding to the odd ones, on the abscissa.[4] The result is a set of curves whose maxima and minima lie alternately at $k = 0$.

The forbidden energy values are directly seen on the ordinate or at $k = \pm\frac{\pi h}{a}$. Since in the general case the expansion coefficients of the potential decrease with the number, the forbidden bands in the spectrum become narrower. Thus, the example of an electron weakly bound with a lattice shows how the general demands imposed on electron states in a periodic field are satisfied.

[4] The expressions near the discontinuities show how much the given point must be shifted to the left to fall within the principal interval.

Electron State in a Three-dimensional Periodic Field. We shall first show how three-dimensional periodic functions are described. Let a three-dimensional potential function revert to its value through translations along three noncoplanar segments, $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$:

$$U(\mathbf{r} + n_1\mathbf{a}_1) = U(\mathbf{r})$$
$$U(\mathbf{r} + n_2\mathbf{a}_2) = U(\mathbf{r})$$
$$U(\mathbf{r} + n_3\mathbf{a}_3) = U(\mathbf{r})$$

Here n_1, n_2, n_3 are integers of any sign; the noncoplanar condition can, as usual, be written as $\mathbf{a}_1(\mathbf{a}_2 \times \mathbf{a}_3) \neq 0$. Segments $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ are essentially the periods of the crystal lattice, that is, the least translations in which the lattice identically superimposes upon itself.

We introduce the following three vectors:

$$\mathbf{b}_1 = \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1(\mathbf{a}_2 \times \mathbf{a}_3)}, \quad \mathbf{b}_2 = \frac{\mathbf{a}_3 \times \mathbf{a}_1}{\mathbf{a}_1(\mathbf{a}_2 \times \mathbf{a}_3)}, \quad \mathbf{b}_3 = \frac{\mathbf{a}_1 \times \mathbf{a}_2}{\mathbf{a}_1(\mathbf{a}_2 \times \mathbf{a}_3)} \tag{42.15}$$

They are called *reciprocal lattice vectors*. From their definition we obtain directly

$$(\mathbf{a}_i \cdot \mathbf{b}_k) = \delta_{ik} \tag{42.16}$$

At $i = k$ a cyclic permutation shows that the numerator and denominator of the expression $(\mathbf{a}_i \cdot \mathbf{b}_k)$ are the same. At $i \neq k$ a cyclic permutation reduces the numerator to zero. If $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ are mutually perpendicular, then $b_i = 1/a_i$.

With the help of the reciprocal lattice vectors the Fourier expansion of the potential energy appears as follows:

$$U(\mathbf{r}) = \mathrm{Re}\left[\sum_{n_1, n_2, n_3} U_{n_1, n_2, n_3} e^{2\pi i(n_1\mathbf{r}\mathbf{b}_1 + n_2\mathbf{r}\mathbf{b}_2 + n_3\mathbf{r}\mathbf{b}_3)}\right] \tag{42.17}$$

From the relationships (42.16) it is apparent that addition of any integral multiple of vectors $\mathbf{a}_1$, $\mathbf{a}_2$, $\mathbf{a}_3$ to the argument $\mathbf{r}$ alters none of the expansion exponents, which proves the periodicity of $U(\mathbf{r})$.

If we introduce a vector with discrete components

$$\mathbf{b} = n_1\mathbf{b}_1 + n_2\mathbf{b}_2 + n_3\mathbf{b}_3 \tag{42.18}$$

the three-dimensional periodic function $U(\mathbf{r})$ takes the form

$$U(\mathbf{r}) = \mathrm{Re}\left(\sum_{\mathbf{b}} U_{\mathbf{b}} e^{2\pi i \mathbf{b}\mathbf{r}}\right) \tag{42.17}$$

Note that the three numbers n_1, n_2, and n_3 should not, together, be equal to zero, since (42.17) refers only to the variable part of the potential. The constant part refers to the work function.

Reasoning in the same way as in the one-dimensional case, we conclude that the wave function of an electron in a three-dimensional periodic lattice has a form similar to (42.6):

$$\psi_{\mathbf{k}} = e^{i\mathbf{kr}/\hbar} u_{\mathbf{k}}(\mathbf{r}) \tag{42.19}$$

Here $u_{\mathbf{k}}(\mathbf{r})$ is a periodic function with the same period as the lattice. The quasi-momentum $\mathbf{k}$ is determined in the following way:

$$\mathbf{k} = 2\pi h (\xi_1 \mathbf{b}_1 + \xi_2 \mathbf{b}_2 + \xi_3 \mathbf{b}_3) \tag{42.20}$$

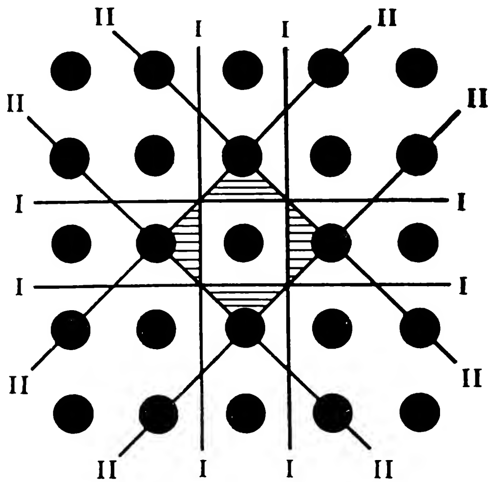


Figure 56

By analogy with the one-dimensional case the numbers ξ can be taken within the limits $-1/2 \leqslant \xi \leqslant +1/2$. Hence in a three-dimensional periodic field there is also a region of quasi-momenta close to the origin of the coordinate system, to which a state with any quasi-momentum reduces. The energy at the boundary of this region undergoes a discontinuity in passing from one band to another.

Let us show, on the example of a simple plane square lattice, how individual zones (called *Brillouin zones*) are constructed. As stated before, in the case of a lattice whose primitive periods are perpendicular, the reciprocal lattice vectors are equal to the reciprocal values of the periods: $b = 1/a$. Let us now construct all the reciprocal lattice vectors from Eq. (42.18), assuming $n_3 = 0$. Denoting the lattice points by solid circles (Figure 56), draw all lines I-I, which divide in half the reciprocal lattice vectors horizontally and

vertically joining the origin with the closest points. On these lines lie the vectors **k**, one of the components of which, either horizontal or vertical, satisfies the condition $k_1 = \pm\pi h b_1$ or $k_2 = \pm\pi h b_2$. The square enclosed by them is the principal Brillouin zone.

The next zone is obtained by drawing lines II-II perpendicular to the reciprocal lattice vectors $\mathbf{b} = \pm\mathbf{b}_1/2 \pm \mathbf{b}_2/2$ passing from the origin to the closest diagonal neighbours. This Brillouin zone is hatched horizontally. The sum of the areas of the resultant triangles

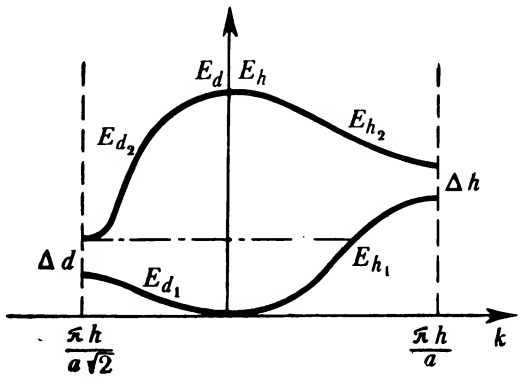


Figure 57

is, as can be seen from Figure 56, equal to the area of the central square. A series of appropriate transfers combines them into a square.

Then in each square, and in all subsequent ones, the energy is a continuous function of the two-dimensional vector **k**, with a discontinuity on the sides of the square.

The boundary of the third zone, not shown in the drawing, corresponds to lines drawn as follows: one pair horizontally at a distance $\pm b$ from the origin, the other pair vertically at a distance $\pm b/2$ from the origin (or vertically at $\pm b/2$ and horizontally at $\pm b$). This yields eight right triangles equal in total area to the principal zone. Then lines are drawn perpendicular to the vectors with components $\pm\mathbf{b}_1$, $\pm\mathbf{b}_2/2$ or $\pm\mathbf{b}_1/2$, $\pm\mathbf{b}_2$, etc.[5]

The location of the energy levels in the two-dimensional case may differ substantially from the one-dimensional case. Imagine cuts of this surface (see Figure 56) in two directions: horizontal and diagonal (Figure 57). We provide them with subscripts h and d, and also 1 and 2 for the first and second zones. Lay off $E_{h_{1,2}}$ to the

[5] All possible combinations of signs are taken.

right of the origin and $E_{d_{1,2}}$ to the left. Then in the two-dimensional case a configuration of the curves is possible (but not necessary) for which the lower end of curve E_{d_2} lies below the upper end of curve E_{h_1}. Thus the energy of the second zone so to say overlaps the first, and no forbidden energy interval occurs on the ordinate. In the one-dimensional case this is impossible. The configuration of energy

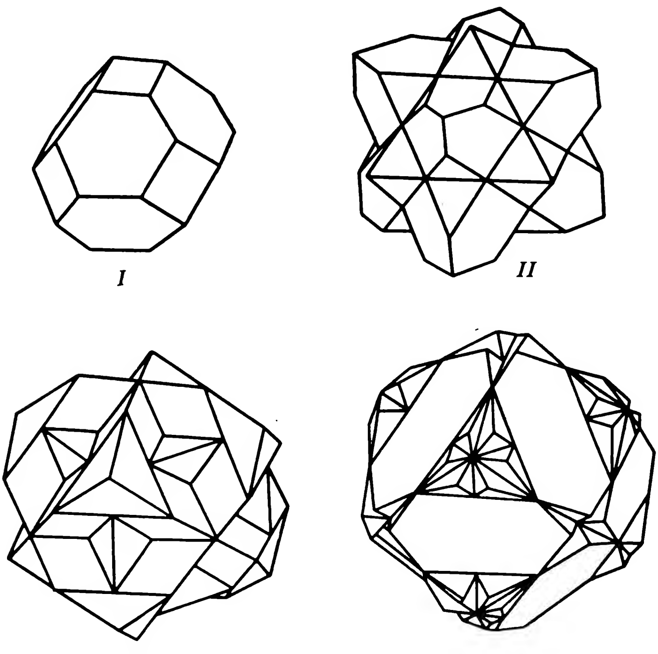


Figure 58

surfaces in two Brillouin zones shown in Figure 57 is called *zone overlap*. A similar configuration is, apparently, possible in the three-dimensional case.

Figure 58 shows the form of the first four zones for a cubic lattice with atoms at the apexes and centres of the faces of the cubes. The drawing offers an idea of the geometric form of zones in three dimensions.

Surfaces of Constant Energy. The motion of any particle can be considered in both geometrical and momentum space. In a periodic field it is natural to employ a quasi-momentum space of vectors **k** defined by Eq. (42.20). Since the energy of an electron in a lattice

is conserved, it is obliged to move along a surface of constant energy $E(\mathbf{k}) = \text{constant}$.

Knowledge of energy bands makes it possible to construct a surface of constant energy. Let us first take a two-dimensional lattice, for which we obtain constant-energy lines instead of surfaces. Imagine them for the case of the rectangular band in Figure 59, and assume that the energy has an extremum at the centre of the band. It is then apparent that there are curves that close around the centre of

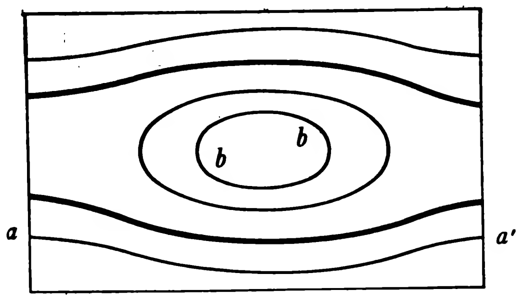


Figure 59

the rectangle, as well as those that approach its boundary, for example at point a or a'. From the construction of the zones, points a and a' are equivalent, so that an electron reaching a passes along the same path again from a'. But the same motion is more conveniently represented in another way: by filling the whole of the **k**-plane with identical rectangles and assuming that on reaching the boundary of a zone the electron passes on into a neighbouring rectangular zone, describing the same path again, etc. Then it passes from a to a' as it were without a discontinuity.

As a result of the described construction we find that there are closed and open paths in the **k**-plane, with constant energy along each of them. The former correspond to finite motions, and the latter to infinite motions.

Similarly, in a **k**-space there are closed and open surfaces of constant energy. For example, if a Brillouin zone has the shape of a right parallelepiped instead of the wavy line in the plane case, we obtain a cylinder with a wavy generatrix (corrugated as it were) and elongated in cross section perpendicular to the axis. The motion of an electron along the generatrix is infinite, and finite at an angle to it. In the next section we shall discuss the manifestations of this in observable effects.

EXERCISES

1. Consider the energy spectrum of an electron in a one-dimensional periodic field having the form of separate rectangular spurs (Figure 60) of height U and width b, and spaced a apart. Assume the energy to be less

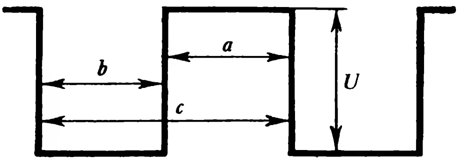


Figure 60

than U. Obtain the answer in the limits Ub = constant as $b \to 0$ (the *Kronig-Penney model*).

Solution. In the region $0 \leqslant x \leqslant a$,

$$\psi = C_1 e^{i\varkappa x} + C_2 e^{-i\varkappa x}, \quad \text{where } \varkappa = \frac{1}{h}(2mE)^{1/2}$$

in the region $b \leqslant x < 0$,

$$\psi = C_3 e^{\lambda x} + C_4 e^{-\lambda x}, \quad \text{where } \lambda = \frac{1}{h}[2m(U-E)]^{1/2}$$

and in the region $a < x \leqslant a + b$,

$$\psi = e^{ikc/h}[C_3 e^{\lambda(x-a-b)} + C_4 e^{-\lambda(x-a-b)}]$$

where, from (42.4) and (42.7), $a + b = c$.

Matching the functions and the derivatives at $x = 0$ and $x = a$, we obtain

$$C_1 + C_2 = C_3 + C_4, \qquad C_1 e^{i\varkappa a} + C_2 e^{-i\varkappa a} = C_3 e^{ikc-\lambda b} + C_4 e^{ikc+\lambda b}$$

$$i\varkappa(C_1 - C_2) = \lambda(C_3 - C_4)$$

$$i\varkappa(C_1 e^{i\varkappa a} - C_2 e^{-i\varkappa a}) = \lambda(C_3 e^{ikc-\lambda b} - C_4 e^{ikc-\lambda b})$$

Excluding C_1 and C_2, we arrive at a set of two homogeneous linear equations for C_3 and C_4, the determinant of which must be zero:

$$\left(1 + \frac{\lambda}{i\varkappa}\right)^2 [e^{2ikc} + 1 - e^{ikc}(e^{-\lambda b - i\varkappa a} + e^{\lambda b + i\varkappa a})]$$
$$- \left(1 - \frac{\lambda}{i\varkappa}\right)^2 [e^{2ikc} + 1 - e^{ikc}(e^{-\lambda b + i\varkappa a} + e^{\lambda b - i\varkappa a})] = 0$$

Multiplying into e^{-ikc}, we arrive at the equation

$$\cos \varkappa a \cosh \lambda b + \frac{\lambda^2 - \varkappa^2}{2\varkappa\lambda} \sin \varkappa a \sinh \lambda b = \cos kc$$

We now perform the limiting process stated in the condition of the problem. Then $\cosh \lambda b = 1$, $\sinh \lambda b = \lambda b$, $a = c$. The product $\lambda^2 b = 2mUb/h^2$. Introducing the notation $mUab/h^2 = A$, we arrive at the equation $A\,(\sin \varkappa a)/(\varkappa a) + \cos \varkappa a = \cos kc$. Its left-hand side is plotted in Figure 61. Its right-hand side lies between the two parallel lines at a distance ± 1 from the abscissa. Hence, the permitted energy intervals lie where

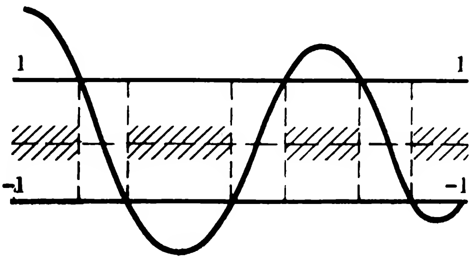


Figure 61

the left-hand side of the equation is greater than 1 or less than -1 (otherwise the curve would not intersect with the lines).

2. Find, in the one-dimensional case, the dependence of E upon k for weakly bound electrons near the edge of the band as a generalization of formula (42.14).

Solution. A nonzero matrix element of the perturbation exists between states with quasi-momenta $k_n + \delta k$ and $k_n - \delta k$ (where k_n corresponds exactly to the edge of the band). To determine the correction to the energy we make use of the method employed in Exercise 3, [Sec. 32], which treats the case of close but not equal energy eigenvalues. The equation for the energy eigenvalue has the form

$$\begin{vmatrix} (k_n + \delta k)^2/(2m) - |E_{0n}| - E & U_n \\ U_n^* & (k_n - \delta k)^2/(2m) - |E_{0n}| - E \end{vmatrix} = 0$$

whence

$$E = \frac{k_n^2}{2m} - |E_{0n}| + \frac{(\delta k)^2}{2m} \pm \left[\frac{k_n^2\,(\delta k)^2}{m^2} + |U_n|^2\right]^{1/2}$$

Consequently

$$v_n = \left(\frac{\partial E}{\partial \delta k}\right)_{\delta k} = 0$$

as was stated.

3. Construct several Brillouin zones of orders higher than the second according to Figure 56.

43

SEMICONDUCTORS AND METALS

Conductivity and the Band Pattern. The band structure of energy levels in crystals makes it possible to understand the principle whereby some crystals conduct electricity while others behave as insulators.

Suppose that at absolute zero all states in all Brillouin zones up to a certain band are occupied by electrons, while in the next and all subsequent bands they are unoccupied. Here, there is no overlap between the last filled and first empty bands, that is, the situation shown in Figure 57 does not occur.

Then a certain energy is required to raise an electron into an empty band. A static electric field, whose energy quantum $\hbar\omega$ is zero, cannot effect this, at least in the case of a weak field, which does not appreciably distort the band pattern.

Since $E(\mathbf{k}) = E(-\mathbf{k})$, in a filled band there are always electrons with quasi-momenta of both signs. If in a static field both these states remain occupied, electric current cannot appear. Therefore a crystal with filled energy bands behaves as a dielectric material.

If a band is only partially filled, then at absolute zero only the lower states are occupied. A continuum of free states adjoins on them directly. In an external field of infinitesimal frequency, that is, a static field, some of the electrons may always pass into neighbouring free states. This produces an excess of electrons with a certain quasi-momentum direction over electrons of opposite quasi-momentum. In other words, current appears. This is the simplest theoretical model of a metal.

Take alkali metals, for example. An electron of an alkali metal is located in the outer s shell, where there are two states corresponding to the possible spin directions. In a crystal lattice the level expands into a band. If, for example, there are n atoms, then in that band there are $2n$ places. At absolute zero, n of the lower places are occupied. The remaining n upper vacancies are unoccupied. Obviously, the boundary between the unoccupied and occupied states in this band is the Fermi level defined in Section 6. The band through which the Fermi boundary passes is called the *conduction band*. When an external field is imposed, the unoccupied states in this band make it possible for the electrons to carry current.

A conduction band can appear in another way. Thus, in beryllium the $2s$ shell is filled. But since the free $2p$ shell is close to that of the $2s$ shell (owing to the small atomic number the levels of beryllium are hydrogenlike), the bands originating from the $2s$ and $2p$

states of atoms in the crystal overlap. Immediately adjoining on the occupied states are unoccupied states, therefore beryllium is a conductor, and not a dielectric substance.

The n-type Semiconductor. Imagine a crystal in which the gap between the occupied and unoccupied bands is of the order of a few hundredths of an electron volt. At room temperature the gap is comparable with the energy of thermal excitation, and some of the electrons pass from the filled band to the free one. In such a state a crystal conducts electricity. The charge is carried both by the electrons that have passed into the initially unoccupied band and by the states remaining unoccupied in the band, which is completely filled at $\theta = 0$ (it is called the *valence band*, because it corresponds to the states of the valence electrons in atoms). The unoccupied states in the valence bands are conventionally called "holes", by analogy with the Dirac holes in negative energy states [Sec. 37]. Holes behave like positively charged electrons.

Unlike the holes in Dirac's theory, semiconductor holes may differ substantially from electrons in the conduction band, not only in the sign of the charge but also, for example, in the dispersion law, that is, in the dependence of energy on quasi-momentum, since they belong to different Brillouin zones.

The properties of semiconductors are most often due not to the transition of electrons from one band to another but to impurities in the substance. If the energy level of an electron in an atom of the impurity is close to the bottom of the conduction band, it is possible for electrons from the impurity to be thermally excited into this band. A certain (usually small) concentration of electrons appears capable of freely moving in the lattice. In this case the semiconductor is said to be of the n-type.

If the impurity has a free level close to the top of the filled band, electrons may be thermally excited into it from below. The remaining holes move as freely in the valence band as electrons in the conduction band. Semiconductors with hole conductivity belong to the p-type.

At absolute zero there are no electrons left in the conduction band and no holes in the valence band, and the semiconductor becomes a dielectric substance. Metals conduct electricity at all temperatures, since free levels adjoin directly on filled levels.

The concentration of electrons or holes is usually small (10^{15} electrons per cubic centimetre), and they therefore move in the lattice as independent particles. Due to low concentration, Pauli's exclusion principle does not affect their state, and the charge carriers in a semiconductor (electrons and holes) form a Boltzmann gas. Thereby the conditions assumed in Lorentz and Drude's classical electron theory of conduction are satisfied.

31—0493

Effective Mass. The close approximation of the theory of semiconductors to the classical electron theory of metals is due to one more reason. In the conduction band electrons occupy states that are close mainly to the bottom of the band. But as can be seen from Exercise 2, Section 42, at $\delta k \ll m \mid U_n \mid / \mid k_n \mid$ the dependence of energy on δk is quadratic:

$$E = \frac{k_n^2}{2m} - \mid E_{0n} \mid \pm \mid U_n \mid + \frac{(\delta k)^2}{2m}\left(1 \pm \frac{k_n^2}{m \mid U_n \mid}\right)$$

$$= \frac{k_n^2}{2m} - \mid E_{0n} \mid \pm \mid U_n \mid + \frac{(\delta k)^2}{2m_*^{\pm}} \qquad (43.1)$$

The energy gap $2 \mid U_n \mid$ between bands is usually much smaller than the energy $k_n^2/(2m)$ at the edge of the band. That is why the sign of the expression in parentheses coincides with that of the root in the initial formula for E. Therefore the E (k) curve has the form of two parabolas: in the upper band the branch goes up: the factor of $(\delta k)^2$ is positive; in the lower band the branch goes down: the factor multiplying $(\delta k)^2$ is negative. But since in a filled band the charge is carried by holes, the factor of $(\delta k)^2$ in the expression for energy is positive, just as the positron mass is positive in Dirac's theory. The factor $m_*^{\pm}$ in the term $(\delta k)^2/(2m_*^{\pm})$ in the expression for energy is called the *effective mass* of the charge carrier. It can be seen from formula (43.1) that it can differ very substantially from the true mass of a free electron. If energy is measured from the edge of a band, its dependence upon quasi-momenta has the form $E = (\delta k)^2/(2m^*)$.

In a real three-dimensional lattice energy depends not only on the absolute value of the quasi-momentum but also on its direction. Since E is a scalar and δk a vector, the general form of the quadratic dependence must be

$$E\,(k) = \frac{1}{2}\, m_{ij}^{-1} \delta k_i \delta k_j \qquad (43.2)$$

where m_{ij}^{-1} is the *tensor of the reciprocal effective mass.* This occurs when the energy minimum of the electrons, or the energy maximum of the holes, is achieved at the middle of the band at $k = 0$. In a crystal with cubic symmetry, m_{ij} (tensor of rank 2) then degenerates into the effective mass scalar m^*. At lower symmetry the dependence remains of the form (43.2).

But in cubic symmetry, too, the energy minimum may not necessarily lie at the middle of a band. For example, silicon has six points in the conduction band where the energy is minimum. These points are displaced in quasi-momentum space from the middle of the band at equal distances along three mutually perpendicular directions of four-fold axes of symmetry. Owing to such a configuration on the axes of symmetry, the surfaces of equal energy

surrounding the minima are ellipsoids of revolution. They are not drawn to scale in Figure 62. The dependence of energy on quasi-momentum near each separate minimum has the form

$$E(k) = \frac{1}{2} m_{ij}^{-1} (k_i - k_{0i})(k_j - k_{0j}) \tag{43.3}$$

Thanks to the symmetry of the ellipsoids the tensor m_{ij} is defined by two numbers.

When the initial state of an atom corresponding to the band in the crystal is not degenerate, there is a simple quadratic dependence

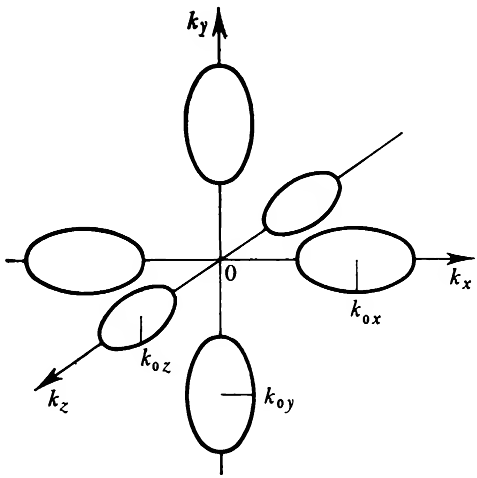


Figure 62

of the energy on the quasi-momentum components. In other cases there are several separate branches expressing the dependence $E(k)$ in the same band. To these branches may approximately correspond charge carriers of different mass: "light" and "heavy" electrons and holes. The valence band of silicon, for example, is degenerate, so that it refers to the hole states.

Extensive elaboration of the theory of n-type semiconductors contributed enormously to the development of semiconductor tech nology. Since the type of semiconductor depends upon the nature o artificially introduced impurities, it is possible to obtain semiconductor materials with planned properties.

Dynamics of an Electron in a Crystal Lattice. If an external field applied to a lattice does not vary appreciably over one period of the lattice, the motion of an electron in the lattice is in many ways like that of a free electron. Assuming the electron to be localized to no greater precision than the confines of one cell, its motion can, by

analogy with (42.9), be represented as the displacement of a wave packet with the velocity $\mathbf{v}_n = \partial E_n/\partial \mathbf{k}$. Here subscript n indicates that the energy belongs to the nth band. As for the quasi-momentum $\mathbf{k}$, its change in unit time is

$$\frac{d\mathbf{k}}{dt} = e\mathbf{E} + \frac{e}{c}\mathbf{v}_n \times \mathbf{H} \tag{43.4}$$

Like (42.9), this equation is obtained if electron motion is imagined as the displacement of a wave packet localized within one cell.

Equation (43.4) can be applied provided the field does not cause transfers between bands. For this the field must not be too strong, it should vary smoothly in space over the length of one cell, and so slowly in time as not to involve frequencies in the Fourier expansion comparable with the ratio of the forbidden band width to Planck's constant.

In other words, the energy in Eq. (43.4) is treated as a smooth function of $\mathbf{k}$ (periodic in $\mathbf{k}$-space), constructed at the end of the preceding section. In two dimensions the dependence of the energy upon $\mathbf{k}$ represents a "parquetry" made up, for example, of rectangles of the shape shown in Figure 59.

Equation (43.1) refers, strictly speaking, to the quasi-momentum operator, which can in principle be constructed with the help of the functions $\psi_{\mathbf{k}} = e^{i\mathbf{kr}/\hbar}u_{\mathbf{k}}(\mathbf{r})$ according to the general rules [Sec. 26]. But in quasi-classical motion (43.1) becomes an equation between quantities. The quasi-classical criterion consists, as always, in that the integral $\int \mathbf{k}\, d\mathbf{r}$ over the domain of motion is large in comparison with the action quantum.

Assume now that the electric field is zero ($\mathbf{E} = 0$), the magnetic field is uniform and constant, and direct the z axis along it ($H_z = H$).

We substitute the velocity, expressed as the derivative of the energy, into Eq. (43.4) and expand the latter in components (omitting the subscript n for brevity) to get

$$\frac{dk_x}{dt} = \frac{e}{c} H \frac{dy}{dt} = \frac{e}{c} H \frac{\partial E}{\partial k_y} \tag{43.5}$$

$$\frac{dk_y}{dt} = -\frac{e}{c} H \frac{dx}{dt} = -\frac{e}{c} H \frac{\partial E}{\partial k_x} \tag{43.6}$$

$$\frac{dk_z}{dt} = 0 \tag{43.7}$$

It follows from these equations that E is an integral of the motion, exactly as in the case of a free electron in a magnetic field. Indeed,

$$\frac{dE}{dt} = \frac{\partial E}{\partial k_x}\frac{dk_x}{dt} + \frac{\partial E}{\partial k_y}\frac{dk_y}{dt} + \frac{\partial E}{\partial k_z}\frac{dk_z}{dt}$$

$$= \frac{e}{c} H \left(\frac{\partial E}{\partial k_x}\frac{\partial E}{\partial k_y} - \frac{\partial E}{\partial k_y}\frac{\partial E}{\partial k_x}\right) = 0 \tag{43.8}$$

Thus, the path of an electron in **k**-space is the intersection of the surface $E(\mathbf{k})$ = constant with the plane k_z. If the surface is closed, the path is closed everywhere. At the end of Section 42 it was pointed out that a surface of constant energy may not necessarily be closed (it can, for example, have the shape of a corrugated cylinder). Then a path lying in any plane in **k**-space and passing through the axis of the cylinder is not closed. Such paths are called "open".

Equations (43.5) and (43.6) show that the scalar product $(d\mathbf{r}/dt) \times \times (d\mathbf{k}/dt)$ is zero, that is, that the velocity in geometric space is everywhere perpendicular to the "velocity" in **k**-space. Thus the paths in these spaces are at right angles.

The Hall Effect. If a conductor carrying an electric current is placed in a magnetic field perpendicular to the current, the current carriers experience a deflecting force in the third perpendicular direction. The force deflects them in the same direction regardless of the sign of the charge of the carriers, because the product ev is of the same sign for all of them. But when an excess of charges forms on one end of the conductor, an electric field perpendicular to the current and the magnetic field appears. This is known as the *Hall effect*. The magnitude of the perpendicular field is

$$E_{\perp} = RjH \tag{43.9}$$

The factor R is called the Hall coefficient.

We shall now show how the Hall coefficient is determined from the transport equation for electrons in a crystal. We write the equation in terms of the corresponding relaxation time, the meaning of which we shall not go into here. Assuming the conductor to be homogeneous, we obtain an equation similar to (41.33):

$$\left(\frac{df}{dt}\right)_{\text{total}} = \frac{\partial f}{\partial \mathbf{k}}\frac{d\mathbf{k}}{dt} + \frac{f - f_0}{\tau} = 0 \tag{43.10}$$

But unlike (41.33), the factor disturbing the statistical equilibrium of the electrons is the Lorentz force applied to them. In (43.10) we substitute the Lorentz force from Eq. (43.4) for $d\mathbf{k}/dt$ to get

$$\left(e\mathbf{E} + \frac{e}{c}\,\mathbf{v} \times \mathbf{H}\right)\frac{\partial f}{\partial \mathbf{k}} + \frac{f - f_0}{\tau} = 0 \tag{43.11}$$

with $\mathbf{v} = \partial E/\partial \mathbf{k}$.

We seek the distribution function in the form of a sum of three terms:

$$f = f_0 + f_1 + f_2 \tag{43.12}$$

Here, f_0 is the equilibrium distribution function, f_1 is proportional to the electric field, and f_2 is proportional to the magnetic field. Substituting this expansion into (43.11) and retaining only the first

nonvanishing terms of the expansions, we obtain

$$eE\,\frac{\partial f_0}{\partial \mathbf{k}}+\frac{e}{c}(\mathbf{v}\times\mathbf{H})\left(\frac{\partial f_0}{\partial \mathbf{k}}+\frac{\partial f_1}{\partial \mathbf{k}}\right)+\frac{f_1+f_2}{\tau}=0 \tag{43.13}$$

But $\partial f_0/\partial\mathbf{k}=(\partial f_0/\partial E)\,(\partial E/\partial\mathbf{k})=(\partial f_0/\partial E)\,\mathbf{v}$. Multiplied by $\mathbf{v}\times\mathbf{H}$ this term yields zero. Therefore, comparing the expressions proportional to $\mathbf{E}$ and $\mathbf{H}$, we obtain two equations

$$f_1=-\tau e\mathbf{E}\,\frac{\partial f_0}{\partial \mathbf{k}} \tag{43.14}$$

$$f_2=-\frac{\tau e}{c}(\mathbf{v}\times\mathbf{H})\,\frac{\partial f_1}{\partial \mathbf{k}} \tag{43.15}$$

Let us now compute the current components. The longitudinal component is computed from the first correction to the distribution sunction (so as to avoid confusion with the relaxation time τ we fhall write $dV_{\mathbf{k}}$ instead of $d\tau_{\mathbf{k}}$)

$$\mathbf{j}^{(1)}=ne\int \mathbf{v}f_1\,dV_{\mathbf{k}}=-n\tau e^2\int\left(\mathbf{E}\cdot\frac{\partial f_0}{\partial \mathbf{k}}\right)\frac{\partial E}{\partial \mathbf{k}}\,dV_{\mathbf{k}}$$

Integrating by parts and taking into account that at the integration limits f_0 becomes zero, we obtain

$$\mathbf{j}^{(1)}=n\tau e^2\int f_0\left(\mathbf{E}\,\frac{\partial}{\partial \mathbf{k}}\right)\frac{\partial E}{\partial \mathbf{k}}\,dV_{\mathbf{k}}$$

But since the mean values of the second derivatives of E with respect to different components of $\mathbf{k}$ are zero, we obtain from this the first term for the current:

$$\mathbf{j}^{(1)}=n\tau e^2\mathbf{E}\int f_0\,\frac{\partial^2 E}{\partial k_x^2}\,dV_{\mathbf{k}} \tag{43.16}$$

(we assume the x axis to be directed along $\mathbf{E}$).

Now we find the second component of the current:

$$\mathbf{j}^{(2)}=ne\int \mathbf{v}f_2\,dV_{\mathbf{k}}=-\frac{n\tau e^2}{c}\int\frac{\partial E}{\partial \mathbf{k}}\left[\left(\frac{\partial E}{\partial \mathbf{k}}\times\mathbf{H}\right)\frac{\partial f_1}{\partial \mathbf{k}}\right]dV_{\mathbf{k}}$$

Integrating once by parts, we obtain

$$\begin{aligned}\mathbf{j}^{(2)}&=\frac{ne^2\tau}{c}\int f_1\left[\frac{\partial}{\partial \mathbf{k}}\left(\frac{\partial E}{\partial \mathbf{k}}\times\mathbf{H}\right)\right]\frac{\partial E}{\partial \mathbf{k}}\,dV_{\mathbf{k}}\\&=\frac{ne^2\tau}{c}\int f_1\left[\left(\frac{\partial E}{\partial \mathbf{k}}\times\mathbf{H}\right)\frac{\partial}{\partial \mathbf{k}}\right]\frac{\partial E}{\partial \mathbf{k}}\,dV_{\mathbf{k}}\end{aligned} \tag{43.17}$$

We did not carry out the differentiation under the sign of the vector product, because curl grad $=0$ for any function. We substitute f_1 into (43.17) and integrate by parts again to get

$$\mathbf{j}^{(2)}=\frac{ne^3\pi^2}{c}\int f_0\left(\mathbf{E}\cdot\frac{\partial}{\partial \mathbf{k}}\right)\left[\left(\frac{\partial E}{\partial \mathbf{k}}\times\mathbf{H}\right)\frac{\partial}{\partial \mathbf{k}}\right]\frac{\partial E}{\partial \mathbf{k}}\,dV_{\mathbf{k}} \tag{43.18}$$

For the integral (43.18) not to vanish the components of **k** with respect to which the differentiation is carried out must be pairwise equal. This is conveniently checked in tensor notation equivalent to (43.18):

$$j_i^{(2)} = \frac{ne^3\tau^2}{c} \int f_0 E_l \frac{\partial}{\partial k_l} \varepsilon_{jmn} \frac{\partial E}{\partial k_j} H_m \frac{\partial^2 E}{\partial k_n\, \partial k_i} dV_{\mathbf{k}} \tag{43.19}$$

Here ε_{jmn} is a totally antisymmetric tensor. The subscripts can coincide in one of two ways: either $n = i$, $j = l$, or $l = n$, $j = i$. Let **H** be directed along the z axis and **E** along the x axis. Then in the terms involving second derivatives the term $\mathbf{j}^{(2)}$ will have only the y-component

$$j_y^{(2)} = \frac{ne^3\tau^2}{c} \int f_0 \left[\varepsilon_{132} \frac{\partial^2 E}{\partial k_x^2} \frac{\partial^2 E}{\partial k_y^2} + \varepsilon_{231} \left(\frac{\partial^2 E}{\partial k_x\, \partial k_y} \right)^2 \right] E_x H_z\, dV_{\mathbf{k}}$$

But since $\varepsilon_{231} = 1$, $\varepsilon_{132} = -1$, and $E_x H_z = -(\mathbf{E} \times \mathbf{H})_y$, we obtain the vector equation

$$\mathbf{j}^{(2)} = \frac{ne^3\tau^2}{c} \left\{ \int f_0 \left[\frac{\partial^2 E}{\partial k_x^2} \frac{\partial^2 E}{\partial k_y^2} - \left(\frac{\partial^2 E}{\partial k_x\, \partial k_y} \right)^2 \right] dV_{\mathbf{k}} \right\} (\mathbf{E} \times \mathbf{H}) \tag{43.20}$$

We shall restrict ourselves to the consideration of nondegenerate bands in semiconductors in which the energy depends upon the quasi-momentum quadratically. Then the terms involving the third derivatives with respect to energy become zero.

Let us now show how to compute the Hall coefficient R from formulas (43.16) and (43.20). Let there be an equation of the form

$$\mathbf{j} = \sigma \mathbf{E} + \lambda \mathbf{E} \times \mathbf{H} \tag{43.21}$$

Assuming the magnetic field to be weak, in the first approximation we substitute $\mathbf{E} = \mathbf{j}/\sigma$ into the vector product. Then in the next approximation we obtain

$$\mathbf{E} = \frac{1}{\sigma} \mathbf{j} - \frac{\lambda}{\sigma^2} \mathbf{j} \times \mathbf{H} \tag{43.22}$$

Comparing this with the definition of the Hall coefficient, we find that

$$R = \frac{\lambda}{\sigma^2} \tag{43.23}$$

Consider the case of an isotropic (scalar) effective mass m_* and write

$$E = \frac{1}{2m_*} (k_x^2 + k_y^2 + k_z^2) \tag{43.24}$$

whence

$$\frac{\partial^2 E}{\partial k_x^2} = \frac{\partial^2 E}{\partial k_y^2} = \frac{1}{m_*}, \qquad \frac{\partial^2 E}{\partial k_x\, \partial k_y} = 0$$

Furthermore, since $\int f_0 \, dV_{\mathbf{k}} = 1$,

$$R = \frac{1}{nec} \tag{43.25}$$

This formula can be used to determine the number of carriers directly if the Hall coefficient R is known. But in such simple form R is expressed in terms of the number of carriers only for the case of semiconductors of the pure n-type or the pure p-type. In the case of mixed conduction, $R = [ec(n_1 - n_2)]^{-1}$. Of course, the applications of this formula are limited, since it assumes that the dependence $E(\mathbf{k})$ for charge carriers of both signs has the form (43.24), in which m_* and τ are assumed the same for both types of carriers.
The sign of R gives the sign of the carriers, that is, the type of semiconductor, provided the carriers are all of the same sign.

Current Carriers in Metals. In most metals the number of electrons is equal, or almost equal, to the number of atoms. The discrepancy may be due to band overlap (Sec. 42), when the band to which a state belongs cannot be defined unambiguously.

In any case, the electron density in metals is high, so that there are no grounds for treating them as a gas of independent particles. Electrons in a metal form a quantum fluid rather than a gas. Another example of a quantum fluid we have encountered before is liquid helium (Sec. 19). The excited states of a quantum liquid resemble individual particles. If such a liquid is homogeneous (liquid helium), the excited states are characterized by energy and momentum. The constant of the motion of the excitations of the electron fluid of a metal in an ion lattice is the quasi-momentum $\mathbf{k}$. Furthermore, unlike excitations in liquid helium II, which behave like bosons, excitations in metals are subject to Fermi statistics and transport charge.

It should not be imagined that all we have is a simple change of names: what was considered a real particle (an electron) is now called an "excitation" or quasi-particle. The Fermi-liquid theory of electrons in a metal also predicts collective effects in which the electrons behave like plasma with excitations quite unlike those of individual electrons. These excitations have been actually observed, but we shall not take this up. Moreover, instead of using the term "excitation" we shall continue to say "electron".

Fermi Surfaces in a Metal. In Section 6 the Fermi distribution was considered in application to a gas whose energy-momentum dependence was given by the simple formula $E = p^2/2m$. Then at absolute zero the particle states fill a Fermi sphere with a boundary energy defined by Eq. (6.6). The degeneracy criteria obtained in Section 6 remain valid for the more complex dependence of energy on the

quasi-momentum of an electron in a metal, but the specific shape of a Fermi surface is occasionally highly contorted.

Construction of such surfaces is examined in *The Electronic Theory of Metals* by I. M. Lifshits *et al.* (Moscow, 1971), and *The Theory of Normal Metals* by A. A. Abrikosov (Moscow, 1972). Here we shall consider only the simplest of the typical Fermi surfaces.

As pointed out before, at absolute zero, that is, in the ground state, the conduction electrons of alkali metals occupy half the corresponding Brillouin zone. In the present case the energy minimum corresponds to the centre of a cube in **k**-space. By virtue of symmetry the lower states consecutively fill spherical surfaces surrounding the centre. Somewhat unexpectedly it develops that the surface surrounding half the lower states in the band comes very close in shape to a sphere: much closer than could be expected after comparing the diameter of such a sphere in **k**-space with the side of the cube. Therefore the spherical surfaces considered in Section 6 in the case of alkali metals describe real relationships.

Any intersection of a sphere with a plane is a closed curve. Therefore the paths of electrons in a magnetic field applied to a metal are always closed curves. There are various ways of experimentally studying such curves, making it possible to "feel" a Fermi surface and determine its shape. A detailed theory of the phenomena whereby the Fermi surfaces of metals are determined was in the main elaborated by I. M. Lifshits and his associates. Here we shall consider one very simple example of determining the general character of a Fermi surface.

Let us take the case represented in Figure 59. If the electron states fill the figure outlined by the solid line, the Fermi surface continues on both sides and has the form of a corrugated cylinder of unlimited length. Such a surface is called *open* (as distinct from the closed Fermi surfaces of alkali metals).

The Fermi surface of gold, silver and copper (Figure 63) is also open. The domain of occupied states in each cell of **k**-space resembles a sphere, but it reaches and adjoins on the six sides of the cube. On each side a near-circular closed line is formed. The domain within the given cell communicates with other such domains of neighbouring cells through the circles thus traced on all six sides. The result is a structure of near-spherical cavities, each with six bridges.

Interactions Between Electrons and Phonons. It was pointed out before that in a perfect crystal lattice an electron travels with constant speed. As can be seen from Eq. (43.4), in a constant electric field the quasi-momentum and hence the energy of an electron increases. Neither is transferred to a perfect lattice, which means that the electron moves without encountering any resistance.

The finite (nonzero) resistance of metals is due to various distortions of their crystal lattice. A lattice may be distorted by impurities, or in a pure substance various defects may have appeared during formation or dislocation. But even in a lattice with no defects there is the chaotic thermal motion of atoms, which disturbs its strict periodicity.

The amplitude of heat displacements is not great in comparison with the lattice period. Even at melting point the atoms of metals

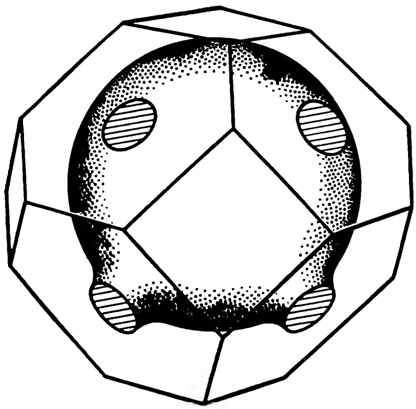


Figure 63

are displaced from their equilibrium positions by no more than one-tenth of the interatomic distance. Therefore the potential energy of an electron in the field of a lattice distorted by thermal motion is represented as a power expansion in the amplitudes of atomic vibrations up to linear terms:

$$U(\mathbf{r}-\mathbf{R}) = U(\mathbf{r}-\mathbf{R}_0) \sum_{\mathbf{n}} \mathbf{a}_{\mathbf{n}} (\operatorname{grad} U)_{\mathbf{R}_{\mathbf{n}}=\mathbf{R}_{0\mathbf{n}}} \tag{43.26}$$

Here $\mathbf{R}$ symbolically denotes the position of all the atoms of the lattice, $\mathbf{R}_0$ is their equilibrium configuration, and $\mathbf{a}_{\mathbf{n}} = \mathbf{R}_{\mathbf{n}} - \mathbf{R}_{0\mathbf{n}}$ (where the subscript $\mathbf{n}$ refers to the number of the lattice point, so that it represents three integers of any sign, n_1, n_2, n_3).

The displacement of the $\mathbf{n}$-th atom is expanded in the normal oscillation modes of the lattice as follows:

$$\mathbf{a}_{\mathbf{n}} = \sum_{\mathbf{f},\sigma} Q_{\mathbf{f}\sigma} e^{i\mathbf{f}\mathbf{n}} \mathbf{e}_{\mathbf{f}\sigma} + \text{complex conjugate} \tag{43.27}$$

Here $Q_{\mathbf{f}\sigma}$ is the normal coordinate of an oscillation having the wave vector $\mathbf{f}$ and polarization σ. This oscillation has the form of a plane wave travelling along discrete atoms whose displacements are in the

direction of the polarization vector $\mathbf{e}_{f\sigma}$. Unlike Section 4, here the wave vector is denoted $\mathbf{f}$ so as not to confuse it with the quasi-momentum of the electron.

Let us express the perturbation energy of the lattice oscillations acting on an electron in terms of the normal oscillation amplitudes:

$$\sum_{\mathbf{n}} \mathbf{a}_{\mathbf{n}} (\mathrm{grad}\, U)_{\mathbf{R}_{\mathbf{n}}=\mathbf{R}_{0\mathbf{n}}}$$

$$= \sum_{\mathbf{f},\sigma} \mathbf{e}_{\mathbf{f}\sigma} Q_{\mathbf{f}\sigma} \sum_{\mathbf{n}} e^{i\mathbf{f}\mathbf{n}} (\mathrm{grad}\, U)_{\mathbf{R}_{\mathbf{n}}=\mathbf{R}_{0\mathbf{n}}} \qquad (43.28)$$

We go over from $\mathbf{n}$ to $n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3 \equiv \mathbf{v}$, and interpret $\mathbf{f}$ as the reciprocal lattice vector: $f_1\mathbf{b}_1 + f_2\mathbf{b}_2 + f_3\mathbf{b}_3$, so as to preserve dimensionality. We represent the internal sum over $\mathbf{n}$ as follows:

$$\sum_{\mathbf{n}} (\ldots) = e^{i\mathbf{f}\mathbf{r}} \sum_{\mathbf{v}} e^{i\mathbf{f}(\mathbf{v}-\mathbf{r})} (\mathrm{grad}\, U)_{\mathbf{R}_0=\mathbf{R}_{0\mathbf{v}}} \qquad (43.29)$$

The factor $e^{i\mathbf{f}\mathbf{r}}$, that is, the sum over $\mathbf{v}$, which we denote Σ_1, now has the period of the lattice. Indeed, a displacement over an integral number of lattice periods along $\mathbf{r}$ means simply a change in the order of summing over $\mathbf{v}$, and $(\mathrm{grad}\, U)_{\mathbf{R}_0=\mathbf{R}_{\mathbf{n}}}$ is a periodic function.

Let us now find the matrix element of the perturbation energy according to the electron states with quasi-momenta $\mathbf{k}$ and $\mathbf{k}'$. This matrix element involves the integral

$$V_{\mathbf{k}\mathbf{k}'\mathbf{f}} \propto \int e^{i(\mathbf{k}-\mathbf{k}'\pm h\mathbf{f})\mathbf{r}/h} u^*_{\mathbf{k}'}(\mathbf{r})\, u_{\mathbf{k}}(\mathbf{r}) \sum\nolimits_1 (\mathbf{r})\, dV \qquad (43.30)$$

Here $u^*_{\mathbf{k}'}$, $u_{\mathbf{k}}$ and Σ_1 are functions that have the same period as the lattice. We transform the expression so that the integration is carried out over one period and is then followed by a summation over all elementary cells of the lattice. Then the following factor comes out from under the integral sign:

$$S(\mathbf{k}, \mathbf{k}', \mathbf{f}) \equiv \sum_{\mathbf{v}} e^{i(\mathbf{k}-\mathbf{k}'\pm h\mathbf{f})\mathbf{v}/h} \qquad (43.31a)$$

The sum along the discrete vector $\mathbf{v}$ resolves into the product of three sums of the form

$$S_1(\mathbf{k}, \mathbf{k}', \mathbf{f}) \equiv \sum_{n_1=-\infty}^{\infty} e^{i(\mathbf{k}-\mathbf{k}'\pm h\mathbf{f})\mathbf{a}_1 n_1/h} \qquad (43.31b)$$

But such a sum does not vanish only if the factor multiplying $\mathbf{a}_1$ is equal to zero or to $2\pi\mathbf{b}_1 h$ (and all the terms are equal to 1). The sum becomes infinite, so that it possesses the property of a δ-function of its argument. Therefore, in the interaction of an electron with the lattice oscillations the projection of vector $\mathbf{k} - \mathbf{k}' \pm h\mathbf{f}$ on $\mathbf{b}_1$ is either zero or it varies by $\pm 2\pi h\mathbf{b}_1^*$. (An integral multiple of $2\pi h\mathbf{b}_1$ is

precluded by the fact that a change in $\mathbf{k}$ greater than by $2\pi h\mathbf{b}_1$ would require a transition to another Brillouin zone, but the distance between bands is usually much greater than the energy of a phonon.) This is also true of the projections of vector $\mathbf{k} - \mathbf{k}' \pm h\mathbf{f}$ on $\mathbf{b}_2$ and $\mathbf{b}_3$. It can therefore be said that in the interaction of an electron with lattice oscillations its quasi-momentum components along the three main principal vectors $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ vary by $h\mathbf{f}_i$ or by $h(\mathbf{f}_i \pm 2\pi\mathbf{b}_1)$ (where $i = 1, 2, 3$). We have what is known as a quasi-conservation law.

Matrix elements of amplitude $Q_{\mathbf{f}\sigma}$ differ from zero only if the quantum number of the normal oscillation, $N_{\mathbf{f}\sigma}$, varies by ± 1 [Secs. 27 and 36]. The matrix element is accordingly proportional to $(N_{\mathbf{f}\sigma} + 1)^{1/2}$ or $(N_{\mathbf{f}\sigma})^{1/2}$.

In such transitions one normal oscillation quantum $h\omega_{\mathbf{f}\sigma}$ is emitted or absorbed, and the energy of the electron varies accordingly by $\pm h\omega_{\mathbf{f}\sigma}$. Therefore two conservation laws hold in the interactions between electrons and lattice oscillations: an exact law for energy and an approximate one (up to an accuracy of $\pm 2\pi h\mathbf{b}_i$) for quasi-momentum.

These processes are conveniently described as the absorption or emission of a phonon, that is, a quasi-particle representing the lattice oscillations. Phonons were already discussed in Section 4. An interaction of the two "gases", the electron gas and the phonon gas, so to say, occurs. But it should be remembered that every phonon exists either before or after colliding with an electron (it is absorbed or emitted).

The Transport Equation for Electrons in a Metal. The probability of an electron colliding with a phonon is proportional to the square of the matrix element of the perturbation energy [32.42]. It therefore involves the factor $N_{\mathbf{f}\sigma} + 1$ or $N_{\mathbf{f}\sigma}$, depending on whether the phonon is emitted or absorbed. As a result of the collision the electron passes from a state with energy E and quasi-momentum $\mathbf{k}$ to a state with energy and quasi-momentum equal respectively to $E \pm h\omega_{\mathbf{f}\sigma}$ and $\mathbf{k} \pm h\mathbf{f}$. For simplicity we have omitted the case when the quantities $2\pi h\mathbf{b}_i$ are added to the quasi-momentum projections on the directions $\mathbf{b}_1$, $\mathbf{b}_2$, $\mathbf{b}_3$. Note that $2\pi h\mathbf{b}_i$ cannot be multiplied by a factor greater than unity, since this would correspond to a displacement of the electron into a neighbouring band. But the distance between bands is a quantity of the order of the usual scale of electron energies (several electron volts), while the energy of a phonon is of the order of the Debye temperature, which does not exceed 0.02 eV, or less.

By Paul's exclusion principle, an electron can transfer only to a state unoccupied by another electron. Therefore the number of transitions in unit time must contain the factor $1 - f(E', \mathbf{k}')$,

which was discussed in Section 4. It should be recalled that the fermion distribution function $f(E, \mathbf{k})$ yields the probability that state $(E', \mathbf{k}')$ is occupied.

Let us now write the balance equation for all transitions from a given state into others and from other states into the given state:

$$\frac{\partial f}{\partial t}+\mathbf{v}\frac{\partial f}{\partial \mathbf{r}}+\left(e\mathbf{E}+\frac{e}{c}\mathbf{v}\times\mathbf{H}\right)\frac{\partial f}{\partial \mathbf{k}}+\sum_{\sigma}\int\frac{dV_f}{(2\pi h)^3}$$
$$\times\{W(\mathbf{k},\mathbf{k}')\,[f(E,\mathbf{k})\,(1-f(E',\mathbf{k}'))\,(N_{\mathbf{f}\sigma}+1)$$
$$-f(E',\mathbf{k}')\,(1-f(E,\mathbf{k}))\,N_{\mathbf{f}\sigma}]$$
$$+W(\mathbf{k},\mathbf{k}'')\,[f(E,\mathbf{k})\,(1-f(E'',\mathbf{k}''))\,N_{\mathbf{f}\sigma}$$
$$-f(E'',\mathbf{k}'')\,(1-f(E,\mathbf{k}))\,(N_{\mathbf{f}\sigma}+1)]\}+\ldots=0 \qquad (43.32)$$

where $E' = E - h\omega_{\mathbf{f}\sigma}$, $\mathbf{k}' = \mathbf{k} - h\mathbf{f}$, $E'' = E + h\omega_{\mathbf{f}\sigma}$, and $\mathbf{k}'' = \mathbf{k} + h\mathbf{f}$.

The dots stand for the collision integrals in which the quasi-momentum components vary by $2\pi h\mathbf{b}_i$.

This rather long equation possesses a simple meaning: it takes into account that fermions may only go over to states unoccupied by other fermions, that the probability of phonon emission and absorption is proportional to $N_{\mathbf{f}\sigma} + 1$ or $N_{\mathbf{f}\sigma}$ respectively, and that a transition of an electron from a given state is possible with either the emission or absorption of a phonon. Accordingly, under the integral sign in (43.32) there are two terms in square brackets involving arguments E', $\mathbf{k}'$ and E'', $\mathbf{k}''$.

If a metal is not situated in an external field ($\mathbf{E} = 0$ and $\mathbf{H} = 0$), is homogeneous ($\partial f/\partial \mathbf{r} = 0$), and in a steady state ($\partial f/\partial t = 0$), then, as in the Boltzmann transport equation (41.21), the collision integral must vanish. This occurs if the Fermi distribution (6.4)

$$f(E)=\frac{1}{\exp[(\mu-E)/\theta]+1}$$

is substituted for $f(E)$, and the Planck distribution (3.7)

$$N_{\mathbf{f}\sigma}=\frac{1}{\exp(h\omega_{\mathbf{f}\sigma})-1}$$

for $N_{\mathbf{f}\sigma}$. This is easily verified by simple computations.

Resistance of Metals at High and Low Temperatures. The temperature dependence of resistance can easily be evaluated in two limiting cases: when $\theta \gg \theta_D$ and when $\theta \ll \theta_D$ (where θ_D is the Debye temperature).

In the first case, when the temperature is high in comparison with the Debye temperature, the energy of a phonon (which is equal

to or less than θ_D) is much smaller than θ. Consequently

$$\exp(h\omega_{f\sigma}) = 1 + h\omega_{f\sigma}/\theta, \qquad N_{f\sigma} = \frac{\theta}{h\omega_{f\sigma}} > \frac{\theta}{\theta_D} \gg 1$$

Neglecting unity in comparison with $N_{f\sigma}$, that is, neglecting spontaneous emission in comparison with stimulated emission, cancels out the characteristic Fermi terms ff' in the transport equation (43.32), leaving an equation of the same type as for particles not subject to Pauli's exclusion principle. In such an equation the collision integral can be substituted with the help of the relaxation time τ, which in this case is inversely proportional to θ (see (43.11)). Then the conductivity can be calculated from Eq. (43.16), in which f_0 should be interpreted as the Fermi distribution function. In the first approximation the integral does not depend on the temperature, because the Fermi distribution is very much like a "step" (Sec. 6). Only the factor τ yields the temperature dependence of the conductivity. That is why at high temperature conductivity is inversely proportional to θ.

At $\theta \ll \theta_D$ the momentum of an electron changes by $h\mathbf{f}$. But at low temperatures only low frequencies of the phonon spectrum ($h\omega_f \lesssim \theta \ll \theta_D$) are excited. For such frequencies, as was shown in Section 4, $|\mathbf{f}| \approx \omega/c$ (where c is the speed of sound).

Consequently, the quasi-momentum of the electron changes by

$$\frac{h\omega}{c} \approx \frac{\theta}{c}$$

Only those electrons undergo transitions whose quasi-momentum lies close to the Fermi surface: deeper inside all states are occupied, and thermally excited electrons have nowhere to go, while outside there are no electrons. Only the smeared region is effective in the distribution. As was pointed out, in metals the quasi-momentum on a Fermi surface is of the same order as near the edge of a Brillouin zone, that is h/a. The relative change in momentum in a transition is

$$\frac{\theta}{c} \Big/ \frac{h}{a} = \frac{\theta a}{hc}$$

But if in the evaluation the velocity of sound is replaced, with the help of formula (4.29), by the maximum frequency, substituting $(V/N')^{1/3}$ for a, then hc/a is replaced by θ_D. Consequently, the change in the quasi-momentum of an electron in a collision with a phonon is but a small fraction of the total quasi-momentum, and the quasi-momentum remains close to the Fermi surface, because the width of the smeared region of the Fermi distribution is also approximately θ. What takes place is as it were a two-dimensional diffusion of the electron's quasi-momentum vector over the Fermi surface. To eval-

uate the diffusion coefficient we should make use of formula (41.13), which holds for random motions of any type. In the present case by "free path" should be understood the change in $\mathbf{k}$ in one collision, that is, $h|\mathbf{f}|$. If $\mathbf{k}$ varies W times per second, then its rate of change is $h|\mathbf{f}|W$. The probability W, as was pointed out, is proportional to the number of phonons N, which in turn varies at low temperature as θ^3. This follows from Exercise 3, Section 4, since at $\theta \ll \theta_D$ the integral yielding the number of phonons extends to infinity.

Thus, the diffusion coefficient of the quasi-momentum of an electron on a Fermi surface, which is equal to the product of the path times the velocity, is proportional to θ^5. The "mobility" of quasi-momenta under the influence of an external electric force deflecting a moving electron has the same temperature dependence. Contrary to the Einstein relation (17.26), the coefficient between both transport coefficients is in this case independent of temperature. This can be understood by taking into account that all dynamic quantities on a Fermi surface are determined by the limiting quasi-momentum independent of temperature. But it is precisely this "mobility" that determines the excess of electrons travelling in the direction of the electric force $e\mathbf{E}$ over electrons travelling in the opposite direction, that is, the ratio between field and current. Consequently, at $\theta \ll \theta_D$ electrical conductivity σ is inversely proportional to θ^{-5}.

In experiments the θ^{-5}-law is not observed for all metals. It is not clear whether this is due to additional electron scattering on impurities and lattice dislocations or on one another (or perhaps even to the inaccuracy of the presented theory of electron-phonon interactions).

Electrical Conductivity of Metals in a Magnetic Field. When experimental data are used to evaluate the free path of electrons in metals, it is found that even at room temperature $l \sim 10^2 a$. An electron manages to travel a long path through a lattice between collisions. In a magnetic field the electron may describe several loops along a closed path, if it corresponds to the closed section of a Fermi surface in $\mathbf{k}$-space.

But this affects the effective value of the electron's mobility. Here we have in mind mobility in conventional space, not on a Fermi surface. Formally, every kind of mobility is evaluated as the square of the path multiplied by the transition probability in unit time, W. If in a magnetic field an electron describes several loops along a length l, its displacement in the plane of the loops is, obviously, not l but only a distance of the order of the radius r_H of the path. The magnetic field is involved in the equations of motion of an electron (43.5)-(43.6) in the product $\mathbf{H} \times d\mathbf{r}$, so that r_H is inversely proportional to H. It is clear from this that for the case of a

closed path an electron's mobility perpendicular to the field is inversely proportional to the square of the field, H^2.

In the item devoted to the Hall effect it was shown that in a magnetic field perpendicular to the electric field a component of the current appears in the third perpendicular direction. To detect it an additional pair of leads must be applied to the conductor, to which an indicator circuit is attached. These leads must assure current take-off from the conductor in a direction perpendicular to the principal direction. Let the latter coincide with the x axis and the magnetic field with the y axis, then the Hall current coincides with the z axis.

But this means that in a magnetic field, the electrical conductivity, which connects the electric field and total current (comprising the conventional and Hall currents) is of tensor nature. If a field **E** along the x axis produces a current along the y axis, then the electrical conductivity is a tensor of rank 2.

It is apparent from formula (43.20) that the off-diagonal components of this tensor change their sign in a permutation of indices. Indeed, if in this formula the electric field is directed once along x and the current along y, and the second time in reverse, for the same H the signs in either case will be opposite.

In the absence of a magnetic field the electrical conductivity tensor reduces to diagonal form with three equal values of σ_0:

$$\begin{pmatrix} \sigma_0 & 0 & 0 \\ 0 & \sigma_0 & 0 \\ 0 & 0 & \sigma_0 \end{pmatrix}$$

In a magnetic field, however, off-diagonal components appear. The general expression for them can be taken from formula (43.9), which in this case yields

$$\sigma_{xy} = -\sigma_{yx} = \frac{1}{RH} \quad \text{since} \quad j_y = \frac{E_x}{RH}, \quad j_x = -\frac{E_y}{RH}$$

The diagonal components of tensor σ in a magnetic field also change. Let us consider separately the case of closed and open paths. They may both occur on the same Fermi surface. Let, for example, the surface be a corrugated cylinder with the axis along z. Let, further, the magnetic field also be acting in the z direction. Then we obtain in **k**-space and in **r**-space closed paths around the cylinder. They lie in the k_x,k_y-plane or the x,y-plane. Accordingly, the path of an electron in a plane perpendicular to the magnetic field is replaced by the path radius r_H, and the diagonal components of the conductivity tensor are multiplied by an additional factor $(r_H/l)^2$ involving H^2 in the denominator.

Consequently, in this case the conductivity tensor in a magnetic field has the form

$$\sigma=\begin{pmatrix}\sigma_0(r_H/l)^2 & (RH)^{-1} & 0\\ -(RH)^{-1} & \sigma_0(r_H/l)^2 & 0\\ 0 & 0 & \sigma_0\end{pmatrix} \tag{43.33}$$

It is interesting to compute the reciprocal tensor, that is, the resistance tensor, with its help. Its components are, as is known, equal to the cofactors of the corresponding components of σ divided by det (σ). In finding det (σ) the term $\sigma_0(r_H/l)^4$, inversely proportional to H^4, must be dropped according to the adopted approximation, since only the first corrections, due to the magnetic field, are being determined. This yields

$$\det(\sigma)=\frac{\sigma_0}{(RH)^2}$$

whence

$$\sigma^{-1}=\begin{pmatrix}\rho & RH & 0\\ -RH & \rho & 0\\ 0 & 0 & 1/\sigma_0\end{pmatrix} \tag{43.34}$$

The diagonal components $(\sigma^{-1})_{xx}$ and $(\sigma^{-1})_{yy}$ are independent of the magnetic field and denoted ρ, but they are not equal to $1/\sigma_0$.

Now let the magnetic field be directed along the y axis, that is, across the axis of the corrugated cylinder. Then the open paths in **k**-space recede along wavy generatrices in the direction of the z axis into infinity. Correspondingly, in coordinate space motion is infinite along the x axis, since from (43.5) the correspondence between paths is achieved by a turn through 90° around the magnetic field. Therefore the free path along the x axis is not shortened and is equal to l, while along the y axis we must continue to take r_H. This yields the conductivity tensor

$$\sigma=\begin{pmatrix}\sigma_0 & 0 & (RH)^{-1}\\ 0 & \sigma_0(r_H/l) & 0\\ -(RH)^{-1} & 0 & \sigma_0\end{pmatrix} \tag{43.35}$$

Again neglecting the terms in the determinant inversely proportional to H^4, we come to the resistance tensor

$$\sigma^{-1}=\begin{pmatrix}1/\sigma_0 & 0 & -(\sigma_0^2RH)^{-1}\\ 0 & (1/\sigma_0)(l/r_H)^2 & 0\\ (\sigma_0^2RH)^{-1} & 0 & 1/\sigma_0\end{pmatrix} \tag{43.36}$$

Along the y axis the resistance increases as the square of the magnetic field. The same occurs at any angle of rotation of the magnetic field

in a plane perpendicular to the cylinder axis. This feature is used as indication that the Fermi surface is open in one direction. The off-diagonal components of the resistance tensor in (43.34) and (43.36) are inversely proportional to the magnetic field.

Superconductivity. In 1911 H. Kamerlingh Onnes found that at several degrees above absolute zero some metals experience a sudden loss of resistance. Subsequently the list of such metals and alloys was greatly enlarged. Numerous signs indicated that the transition to the superconductive state was due to interactions between electrons, it being a typical phase transition. In the absence of a magnetic field it is a phase transition of the second kind, since it is accompanied only by a discontinuity in specific heat; in a magnetic field this phase transition is accompanied by evolution of heat, that is, it is a phase transition of the first kind. But the nature of the interaction responsible for such a combined effect remained a riddle for a long time.

Towards the late 1940's, when techniques were devised for separating macroscopic quantities of metal isotopes, it was found that the transition temperature for some superconductors is inversely proportional to the square root of the atomic weight of the corresponding isotope. This is precisely the same dependence as that of the oscillation frequency of a lattice on the atomic mass (as, incidentally, is the case for any harmonic oscillator). This suggests the involvement of phonons. In 1950 H. Fröhlich showed that the electrons of a metal are capable of interacting by means of phonons. One electron emits a phonon while another absorbs it, which is similar to the way interaction through phonons occurs in vacuum.

Before Fröhlich's work it was considered that electrons were capable only of the Coulomb repulsion. Interaction through phonons, however, was found to involve attraction; it is thus capable of assembling electrons. But this was not yet an explanation of superconductivity. It was found in 1956 by J. Bardeen, L.N. Cooper and J.R. Schrieffer. Their theory was soon improved upon by N.N. Bogoliubov and L.P. Gor'kov.

The gist of the explanation is that the Fröhlich interaction results in a peculiar assembly of electrons in pairs, the binding energy being of the order of the transition temperature into the superconducting state. Then in addition to the electron, or fermion, excitation branch another excitation branch appears in metals similar to the excitation in liquid helium.

Excitations from paired electrons have spin zero and are similar to bosons. At least, they are capable of accumulating in the ground state and forming the corresponding collective superconductor state.

The phenomenon of superconductivity is similar to superfluidity: the ground state due to collective interaction is not destroyed by

separate small disturbances of the order of thermal disturbances. Accordingly, a separate electron is not scattered on lattice phonons, since this would require a change of state of all current-carrying electrons.

Direct experiments carried out after the mentioned theoretical works confirmed that the current carriers in superconductors have double electron charge, so that electron pairing is an actual fact.

With the solution of the superconductivity problem there is no natural phenomenon on the atomic and molecular level which cannot be explained in terms of nonrelativistic quantum mechanics. The unsolved problems of physics lie in the domain of the atomic nucleus and elementary interactions, or of elementary particles.

We may note that quantitatively the microscopic theory of superconductivity is better developed than the theory of superfluidity.

EXERCISES

1. Determine the Hall coefficient for a semiconductor of cubic crystal symmetry but with the energy surfaces located not at the centre of the band (Figure 62).

Solution. The general expressions (43.16) and (43.20) must be averaged over all energy minima lying in the conduction band. The integrand in (43.16) yields

$$\frac{1}{3}\left(\frac{\partial^2 E}{\partial k_x^2}+\frac{\partial^2 E}{\partial k_y^2}+\frac{\partial^2 E}{\partial k_z^2}\right)$$

If the energy is reduced to the principal axes, the second derivatives are respectively equal to $1/m_1$, $1/m_2$, and $1/m_3$. The expression involved in (43.20) is a minor formed from the matrix $\dfrac{\partial^2 E}{\partial k_i \partial k_j}$. In the system of principal axes there remain the minors

$$\frac{1}{m_1 m_2}, \qquad \frac{1}{m_1 m_3}, \qquad \frac{1}{m_2 m_3}$$

Whence, formula (43.25) involves the factor

$$\frac{1}{3}\left(\frac{1}{m_1 m_2}+\frac{1}{m_1 m_3}+\frac{1}{m_2 m_3}\right)\Big/\frac{1}{9}\left(\frac{1}{m_1}+\frac{1}{m_2}+\frac{1}{m_3}\right)^2$$

If the ellipsoids of constant energy possess a symmetry axis, then two effective masses are equal, which is as it should be in this case. Let, for example, $m_1 = m_2 \neq m_3$. For germanium $m_3 \gg m_1 = m_2$. Then the expression written above is equal to 3/4.

2. Find the general formula for the rotation period of an electron along a surface of constant energy in a magnetic field.

Solution. The required period is found as follows:

$$t=\int \frac{dk_x}{dk_x/dt}=\int \frac{dk_x}{(e/c)\,H\,(\partial E/\partial k_y)}=\frac{c}{eH}\int \frac{\partial k_y}{\partial E}\,dk_x$$
$$=\frac{c}{eH}\frac{\partial}{\partial E}\int k_y\,dk_x$$

The derivative with respect to E is taken outside the integral sign because the integration is performed at E = constant. The integral is computed over a closed path in the k_x,k_y-plane and is equal to the cross-sectional area of the surface of constant energy.

3. The surface of constant energy of an electron in a semiconductor has the shape of an ellipsoid $2E = k_1^2/m_1 + k_2^2/m_2 + k_3^2/m_3$. The magnetic field forms angles with the ellipsoid's axes whose cosines are α_1, α_2, and α_3. Determine the rotation period of the electron in a magnetic field.

Solution. The equation of the plane of orbit in **k**-space is

$$\sum_{i=1}^{3} \alpha_i k_i = k$$

We carry out the transformation $k_i \equiv x_i\,(m_i)^{1/2}$. In terms of the x_i's the equation of the plane has the form

$$\sum_i \alpha_i x_i\,(m_i)^{1/2}=k$$

and the surface of constant energy becomes a sphere ·

$$\sum_i x_i^2=2E$$

The length of the perpendicular from the origin to the plane is $k\left(\sum \alpha_i^2 m_i\right)^{-1/2}$. The square of the radius of the circle formed by the intersection of the sphere and the plane is found from the formula

$$r^2=2E-k^2\left(\sum_i \alpha_i^2 m_i\right)^{-1}$$

The volume of a cone whose apex is at the origin of the coordinate system and whose base is this circle is

$$\frac{\pi r^2}{3}k\left(\sum_i \alpha_i^2 m_i\right)^{-1/2}$$

To return to the initial coordinates k_1, k_2, k_3 the volume of the cone must be multiplied by $(m_1m_2m_3)^{1/2}$. We then obtain

$$\frac{\pi}{3}k\,(m_1m_2m_3)^{1/2}\left(\sum_i \alpha_i^2 m_i\right)^{-1/2}\left[2E-k^2\left(\sum_i \alpha_i^2 m_i\right)^{-1}\right]$$

The area of the base of the cone equals its volume divided by one-third of its height, that is, by $k/3$:

$$\pi\,(m_1m_2m_3)^{1/2}\left(\sum_i \alpha_i^2 m_i\right)^{-1/2}\left[2E-k^2\left(\sum_i \alpha_i^2 m_i\right)^{-1}\right]$$

Whence, differentiating with respect to the energy E, we find the rotation period:

$$T=\frac{2\pi c}{eH}\left(\frac{\alpha_1^2}{m_2m_3}+\frac{\alpha_2^2}{m_1m_3}+\frac{\alpha_3^2}{m_2m_3}\right)^{-1/2}$$

The rotation frequency is

$$\omega=\frac{eH}{c}\left(\frac{\alpha_1^2}{m_2m_3}+\frac{\alpha_2^2}{m_1m_3}+\frac{\alpha_3^2}{m_2m_3}\right)^{1/2}$$

In a high-frequency field of the same frequency ω the so-called *cyclotron resonance* is observed. By changing the direction of the magnetic field it is possible to determine the principal values of the mass tensor.

INDEX

TO THE READER

Mir Publishers welcome your comments on the content, translation, and design of this book.

We would also be pleased to receive any proposals you care to make about our future publications.

Our address is:
USSR, 129820, Moscow I-110, GSP
Pervy Rizhsky Pereulok, 2
Mir Publishers

Printed in the Union of Soviet Socialist Republics

NEW BOOKS FROM MIR PUBLISHERS

Mechanics

by S. Strelkov, D.Sc.

This general course of mechanics by the late professor S. Strelkov is meant for students of physical and mathematical departments of universities and teachers colleges. It is based on the material of the seminars and lectures given by the author for many years at the physical department of Moscow State University.

Contents. **Mechanics of Rigid Bodies.** Kinematics of a Particle. Fundamental Laws of Dynamics. Momentum of a System of Bodies. Work and Energy. Friction. Relative Motion. Motion of a Rigid Body. Rolling Friction. Gravitational Attraction of Bodies. **Mechanics of Deformable Bodies.** Mechanics of Deformable Solids. Equilibrium of Fluids and Gases. Motion of Fluids and Gases. Action of a Fluid or Gas Flow on a Body. **Oscillation and Waves. Elements of Acoustics. Fundamentals of Special Theory of Relativity.** Oscillations. Vibrations in a Continuous Medium. Elements of Acoustics. Fundamentals of Special Theory of Relativity. Name Index. Subject Index.

The Theory of Functions of a Complex Variable

by A. Sveshnikov, D.Sc. and A. Tikhonov, Mem. USSR Acad. Sci.

This textbook is intended for students of physico-mathematical departments of colleges and universities; can be used as a reference book by postgraduate students and research workers.

Contents. The Complex Variable and Functions of a Complex Variable. Series of Analytic Functions. Analytic Continuation. Elementary Functions of a Complex Variable. The Laurent Series and Isolated Singular Points. Residues and Their Applications. Conformal Mapping. Analytic Functions in the Solution of Boundary-Value Problems. Fundamentals of Operational Calculus. Saddle-Point Method. The Wiener-Hopf Method. Functions of Many Complex Variables. Appendix: Watson's Method. References. Name Index. Subject Index.

VOLUME 2 OF THIS *COURSE OF THEORETICAL PHYSICS* DEALS WITH STATISTICAL LAWS. THE BASIC STRUCTURE REMAINS ESSENTIALLY THE SAME. THE AUTHOR HAS SELECTED THOSE TOPICS HE FELT TO BE OF GENERAL INTEREST. THE BOOK INCLUDES, FOR INSTANCE, SECTIONS ON FLUCTUATIONS, GIBBS STATISTICS, DETONATION WAVES, FERROMAGNETISM, AND THE THEORY OF SEMICONDUCTORS. *STATISTICAL LAWS* CAN BE READ BY A STUDENT WHO HAS HAD COURSES IN CLASSICAL MECHANICS, ELECTRODYNAMICS, AND QUANTUM MECHANICS. NUMEROUS EXERCISES COMBINE WITH THE MASTERLY COVERAGE OF THE SUBJECT TO MAKE *STATISTICAL LAWS* AN ESSENTIAL TEXT FOR UNIVERSITY AND COLLEGE STUDENTS.

Jacket design by L. Muratova

ALEXANDER S. KOMPANEYETS
(1914-1974)

PROFESSOR ALEXANDER SOLOMONOVICH KOMPANEYETS WAS A LEADING SOVIET THEORETICAL PHYSICIST. FROM 1946 UNTIL HIS UNTIMELY DEATH HE WORKED AT THE INSTITUTE OF CHEMICAL PHYSICS OF THE USSR ACADEMY OF SCIENCES, CONTRIBUTING, AMONG OTHER THINGS, TO THE DEVELOPMENT OF NUCLEAR ENERGY IN THE SOVIET UNION IN ALL ITS ASPECTS.

MIR PUBLISHERS · MOSCOW

Titles of Related Interest from Mir Publishers

PROBLEMS IN THEORETICAL PHYSICS

by L. G. Grechko, V. I. Sugakov, O. F. Tomasevich, and A. M. Fedorchenko

This book is a collection of problems covering mechanics, electrodynamics, nonrelativistic quantum mechanics, statistical physics and thermodynamics. Each section opens with a brief outline of the main laws and relationships used to solve the problems. Also information about the needed mathematical apparatus is included. Along with answers there are guides to solving the most complicated problems. SI units are used throughout the book. *Problems in Theoretical Physics* is intended for physics majors at universities and other institutions of higher learning. Some of the problems are specifically for students majoring in theoretical physics. Certain ones can be used in the physics and mathematics departments of teachers' colleges.
448 pages

ELEMENTS OF APPLIED MATHEMATICS

by Ya. B. Zeldovich and A. D. Myškis

The text describes useful methods of calculation and gives the fundamentals of complex variables, linear differential equations, vectors and vector fields, and the calculus of variations. Formal proofs are largely replaced by leading questions and pointers, thereby achieving simplicity and clarity of exposition. Certain physical problems, in particular those relating to optics, mechanics, and the theory of probability, are analyzed in detail. This book will be of interest to university students as a supplement to their standard textbooks, and to engineers, physicists, and anyone else wishing to brush up on the elements of modern applied mathematics.

656 pages

INTRODUCTION TO PLASMA PHYSICS

by B. M. Smirnov

This book is intended for senior college and graduate students interested in the physics of weakly ionized gas. It deals with the main concepts of the physics of weakly ionized plasmas and describes the characteristics of a practically realizable plasma. The book examines the properties of a weakly ionized gas, the propagation of radiation in such gases, and the elementary interactions between radiation and gas. As an example of a specific plasma it considers the properties of the earth's upper atmosphere.
174 pages

MIR PUBLISHERS · MOSCOW

KOMPANEYETS
A COURSE OF THEORETICAL PHYSICS
2

KOMPANEYETS

A COURSE OF THEORETICAL PHYSICS

2